

3-23-2017

Survival Analysis of US Air Force Officer Retention Rate

Courtney N. Franzen

Follow this and additional works at: <https://scholar.afit.edu/etd>



Part of the [Operational Research Commons](#)

Recommended Citation

Franzen, Courtney N., "Survival Analysis of US Air Force Officer Retention Rate" (2017). *Theses and Dissertations*. 795.
<https://scholar.afit.edu/etd/795>

This Thesis is brought to you for free and open access by AFIT Scholar. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of AFIT Scholar. For more information, please contact richard.mansfield@afit.edu.



SURVIVAL ANALYSIS OF US AIR FORCE RATED OFFICER RETENTION

THESIS

Courtney N. Franzen, Captain, USAF

AFIT-ENS-MS-17-M-129

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views expressed in this document are those of the author and do not reflect the official policy or position of the United States Air Force, the United States Department of Defense or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

SURVIVAL ANALYSIS OF US AIR FORCE RATED OFFICER RETENTION

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the
Degree of Master of Science in Operations Research

Courtney N. Franzen, B.S., M.A.

Captain, USAF

MARCH 2017

DISTRIBUTION STATEMENT A

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

SURVIVAL ANALYSIS OF US AIR FORCE RATED OFFICER RETENTION

THESIS

Courtney N. Franzen, B.S., M.A.
Captain, USAF

Committee Membership:

Dr. Raymond R. Hill
Chair

Lt Col Matthew J. Robbins, PhD
Member

Abstract

Personnel retention is a matter of great interest in the private and public sectors. In the public sector--specifically the US Air Force--maintaining appropriate retention levels of rated officers is paramount as these officers are the backbone of the Air Force's mission. As part of the effort to ensure proper retention rates for rated officers, retention models are created by the Air Force Personnel division that assist in predicting future retention patterns and accession needs. The techniques for creating these models, known as the "sustainment line," involve utilizing average retention percentages obtained from historical data. In this study, more statistical-based methods involving logistic regression analysis and survival analysis are utilized to obtain similar retention models for rated officers. The survival analysis curve produces similar results to the sustainment line, but the sustainment line currently employed is a one-dimensional view of retention patterns. It simply models the rate at which officers leave. The value of the survival curve created in this study is that it can be updated very quickly, is flexible in its construction, and can incorporate covariates into the model that are significant to retention rates. The Air Force has long known that there are external (e.g., economic) factors that impact retention. Using a survival analysis regression model instead of simply modeling the rate at which officers leave, this study was able to identify six demographic and one economic factor that may be significant to rated officer retention. This ultimately could lead to the creation of models that reflect the retention behavior of certain subtypes of officer and give insight that could be used to tailor retention and accession programs so that they are more resource-effective.

Acknowledgements

First, and foremost, I would like to express my deepest gratitude to my advisor, Dr. Raymond R. Hill. He showed endless dedication and patience in guiding me through every step of the thesis process. Additionally, I would like to thank Lt Col Matthew Robbins, Ph.D., for giving his time to review and fine-tune my work. Finally, I would like to thank HAF/A1 and Mr. John Sanzone for providing me the data and support that were crucial to the completion of my thesis.

Courtney N. Franzen

Table of Contents

Abstract	iv
Acknowledgements	v
List of Figures	viii
List of Tables	x
I. Introduction	1
1.1 Problem Background	1
1.2 Research Scope	5
1.3 Issues, Needs, and Limitations	6
1.4 Thesis Outline	9
II. Literature Review and Background	10
2.1 Introduction	10
2.2 Antecedent Studies	10
2.3 Modeling Techniques	13
2.4 Significant Factors	19
2.5 Logistic Regression	23
2.6 Sustainment Line	25
2.7 Survival Analysis	27
III. Data Sources	30
3.1 Introduction	30
3.2 Demographic Data	30
3.3 Economic Data	32
3.4 Political Data	34
IV. Models of Rated Officer Retention	36
4.1 Introduction	36
4.2 Logistic Regression	36
4.2.1. Introduction	36
4.2.2. Data	36
4.2.3. Analysis at Career Field Level	38

4.2.4. Analysis of ABM CYOS Subset p-Values	45
4.2.5. Analysis of CSO CYOS Subset p-Values	46
4.2.6. Analysis of Pilot CYOS Subset p-Values.....	47
4.3 Survival Analysis	48
4.3.1. Introduction	48
4.3.2. Data.....	49
4.3.3. Kaplan-Meier Survival Functions	50
4.3.4. Cox Proportional Hazards Regression.....	56
V. Results and Analysis	61
5.1 Introduction	61
5.2 Results	61
5.3 Analysis of Model Adequacy.....	64
5.4 Analysis of Economic and Political Factors	69
VI. Conclusion	75
6.1 Introduction	75
6.2 Limitations of Work.....	77
6.3 Follow-On Research.....	78
Appendix A. SAS Code for Logistic Regression.....	79
Appendix B. SAS Code for Survival Analysis	83
Appendix C. Kaplan-Meier Curves	89
Bibliography	93

List of Figures

Figure 1. Career Field Health Chart (All Officers).....	26
Figure 2. Example KM Curve [39]	28
Figure 3. Odds Ratio of Retention by Marital Status.....	40
Figure 4. Odds Ratio of Retention by Gender	41
Figure 5. Odds Ratio of Retention by Commissioning Source.....	42
Figure 6. Odds Ratio of Retention by Prior Enlisted Service (CYOS).....	43
Figure 7. Odds Ratio of Retention for Prior Enlisted Service (TAFMS)	43
Figure 8. Odds Ratio of Retention for Distinguished Graduates	44
Figure 9. Odds Ratio of Retention for Officers with Dependents	45
Figure 10. Kaplan Meier Curve for Marital Status (Pilot).....	51
Figure 11. Kaplan-Meier Curve for Gender	52
Figure 12. Kaplan-Meier Curve for Commissioning Source (ABM)	53
Figure 13. Kaplan-Meier Curve for Prior Service (ABM)	54
Figure 14. Kaplan-Meier Curve for DG	55
Figure 15. Kaplan-Meier Curve for Dependents (CSO).....	56
Figure 16. Range of Survival Functions	59
Figure 17. HAF/A1 Sustainment Line for Pilot with KM Curve Overlay.....	62
Figure 18. HAF/A1 Sustainment Line for CSO with KM Curve Overlay	63
Figure 19. HAF/A1 Sustainment Line for ABM with KM Curve Overlay	64
Figure 20. Martingale Residual Plot for Cox Proportional Hazards Model (Pilot)	65
Figure 21. Deviance Residuals Plot for Cox Proportional Hazards Model (Pilot).....	66
Figure 22. Martingale Residual Plot for Cox Proportional Hazards Model (CSO).....	67
Figure 23. Deviance Residual Plot for Cox Proportional Hazards Model (CSO)	67
Figure 24. Martingale Residual Plot for Cox Proportional Hazards Model (ABM)	68
Figure 25. Deviance Residual Plot for Cox Proportional Hazards Model (ABM).....	69
Figure 26. Survival Curve for Pilot and CSOs with Low, Average, and High Female Percentage	71
Figure 27. Survival Curve for Rated Officers with Low, Average, and High New Orders Values	73
Figure 28. KM Curve for Marital Status (CSO)	89
Figure 29. KM Curve for Marital Status (ABM).....	89
Figure 30. KM Curve for Commissioning Source (Pilot).....	90
Figure 31. KM Curve for Commissioning Source (CSO)	90
Figure 32. KM Curve for Prior Service (Pilot).....	91

Figure 33. KM Curve for Prior Service (CSO).....	91
Figure 34. KM Curve for Dependents (Pilot).....	92
Figure 35. KM Curve for Dependents (ABM).....	92

List of Tables

Table 1. Logistic Regression Covariate p-Values for Pilot, CSO, and ABM.....	39
Table 2. p-Values for ABM by CYOS.....	45
Table 3. p-Values for CSO by CYOS.....	46
Table 4. p-Values for Pilot by CYOS	47
Table 5. Cox PHREG Covariate p-Values for Pilot, CSO, and ABM.....	57
Table 6. Cox PHREG Model Parameter Estimates	58
Table 7. Stepwise PHREG Model p-Values	70

I. Introduction

1.1 Problem Background

In an occasional paper written for the Office of the Secretary of Defense (OSD), Asch and Hosek [1] discussed military manpower and personnel policy challenges in the dawn of a transformation effort to redefine the military so that it has the capabilities to overcome all possible future threats. Asch and Hosek pointed out that since the dismantling of conscription in the 1970s, sustaining adequate manpower (a key to mission effectiveness) in an All-Volunteer Force (AVF) has seen many challenges. A consistent theme has been an inability to maintain desired manning levels for certain, important career fields, such as the rated officer career fields.

The Air Force officer corps is broken up into the various career fields necessary to run an independent, self-sufficient organization. These career fields are grouped into categories based upon their relationship to the mission. For example, there are medical officers (e.g., doctors, nurses, dentists, etc.) and there are support officers (e.g., logistics, finance, human resource, etc.). The category of officers encompassing the aviation service career fields – pilots, remotely piloted aircraft (RPA) pilots, combat systems officers (CSO), navigators, and air battle managers (ABM) – are classified as “rated officers” [2]. This terminology is based upon their career field’s requirement to possess aeronautical ratings (e.g., basic CSO rating) to perform specific aviation duties. These ratings require extensive initial training and continuous qualification checks and upgrade training. Unlike the non-rated career fields, having these ratings is crucial to the officer’s ability to execute

their duties. Ratings can be revoked for various reasons (e.g., medical disqualification, disciplinary issues, aviation mishaps, etc.). If an officer's ratings are revoked, then that officer is no longer qualified to perform the duties associated with their career field.

The training necessary to produce rated officers is expensive and varies within and among the rated specialties. For example, pilot training is notorious for the amount of time and financial resources that the government must invest to train a "basic pilot." A U.S. General Accounting Office study estimated the figure for basic pilot training at approximately \$1 million. To fully train a pilot takes approximately two years and is estimated at \$9 million (depending upon the type of aircraft the pilot is selected to fly) [3]. As a rated officer's years of service (YOS) increase, their level of experience, and equivalently, their level of aeronautical ratings increases. If an experienced rated officer is lost due to voluntary separation from the Air Force, the introduction of a new rated officer is not an equivalent exchange due to a lack of experience and/or training. "Indeed, the Air Force would prefer to have [rated officers] choose to stay in the service rather than voluntarily exit. This is because turnover is costly to the Air Force as it is for any firm" [4]. This exact point is echoed in Air Force Instruction (AFI) 11-412, *Flying Operations*, which has an entire chapter dedicated to retention of rated officers because "retaining aircrews the AF has trained and experienced is much more effective, efficient, and responsive than replacing them, since replacement requires substantial time and money to access, train, and experience new aircrews" [2]. Therefore, the unintended loss of an experienced rated officer is not only detrimental to mission effectiveness but will result in the loss of time and financial resources that were invested in a rated officer with the expectation of reaping the benefits of that training over a full, 20-year career.

The desirable outcome for military personnel policies is to recruit and train the exact number of rated officers required to perform the mission over some period of time. However, this is not always feasible due to the voluntary and involuntary separation of officers. Involuntary separation is the result of personnel policies that produce an excess number of personnel to accomplish a mission. For example, recruitment increased with the advent of the Iraq War in 2003. With the drawdown of troops in Iraq and Afghanistan in the subsequent years, the number of missions and the needs of the U.S. military shifted and decreased, which also shifted and decreased the corresponding personnel requirements. This shift resulted in excess personnel levels that necessitated force management (FM) efforts in the Air Force in 2005, 2009, and 2014. These excessive personnel levels were, in part, due to the Air Force's primary assumption that current end strength numbers will not change for the next 30 years. Thus, when the Air Force needed more personnel for the Iraq War, they projected needing these numbers for the next 30 years and recruited accordingly [2]. When that assumption proved false, the FM programs were initiated. The program first encouraged voluntary separation of interested members by providing monetary incentives, and when voluntary separation did not reduce personnel to the desired manning numbers, involuntary separations were conducted.

While the military has infrequently been required to initiate programs to downsize its force (i.e., reduce manpower), in most cases it suffers manpower shortages, especially in certain career fields, such as the rated officer career fields. The primary reason for these shortages are voluntary separations at the completion of active duty service commitments (ADSC). ADSCs are contractual obligations for military members to serve the U.S. military for a specified length of time. The first ADSC acquired by an officer is an

accession ADSC, acquired on the date that an officer is commissioned and entered into active duty. This accession ADSC is traditionally four years with the exception of officers from the United States Air Force Academy (USAFA), which are obligated to five years. Multiple ADSCs can be served concurrently – i.e., multiple ADSCs do not compound onto each other – and rated officers typically accrue at least two ADSCs at the beginning of their career: one for accession and one for training. When a rated officer completes their training (e.g., Undergraduate CSO Training (UCT) for navigators and CSOs), they acquire an additional ADSC for their rated training. Pilots are obligated to serve ten years (concurrent with the four-year obligation for their accession ADSC) after completing their training. For navigators and CSOs, it is six years. For ABMs, it is three years [5]. Additional ADSCs can also be acquired at any point in an officer's career in exchange for specialized or upgrade training, education, promotions, permanent changes of station (PCS), monetary bonuses, etc. [5].

All these opportunities for an officer to acquire an ADSC are also opportunities in which the Air Force can ensure retention of its officers. Once these obligations are served to their completion, however, officers are eligible to voluntarily separate from the Air Force when they choose. This is when manpower retention is most tenuous and when economic and demographic factors have the most influence on an officer's actions. This is also when being able to accurately predict the probability of retention is invaluable. This research seeks to identify the most significant factors influencing retention decisions at these decision points in a rated officer's career and use these factors to create a flexible, updateable model to help predict or anticipate future retention rates of rated officers.

1.2 Research Scope

While a few studies have investigated retention rates for the entire pilot career field, a majority of military manpower studies focus on fighter pilots. Recently, there has been a shift in emphasis. The shortage of RPA pilots has become so prolific that it is reflected in news media alongside the deficiency in fighter pilots [6]; however, studies investigating RPA pilot retention are sparse, likely due to the relatively recent adoption of RPA into military strategy and operations. This study expands officer retention research to all rated career fields by examining the demographic, economic, and political factors most likely to contribute to the voluntary attrition of rated officers. These significant factors are identified using logistic regression. Furthermore, survival analysis is employed to build a mathematical model capable of predicting retention rates for rated officers at given moments in their career.

The research involves five phases. Phase one examines a representative cross-section of the research conducted and the findings reported over the past 15 years regarding military officer retention. Phase two examines the Air Force's current process of determining end strength and therefore defining accession numbers. Phase three analyzes the possible relationships between specific factors and the retention rates of those populations. Phase four utilizes the factors identified as potentially significant to create a model for predicting future attrition rates of certain subgroups. Finally, phase five explores the impact that potential shifts in economic and political factors could have on retention decisions and its potential use to predict future retention behavior based on these fluctuations.

1.3 Issues, Needs, and Limitations

Air Force Specialty Codes (AFSC) are specific alphanumeric designators used to identify officer specialties and level of experience throughout an officer's career. While officers are assigned a specific career field to which they typically remain allocated to throughout their career, the Air Force routinely employs officers outside of their career field in "career broadening" positions. However, when officers are operating in this career broadening environment, they are still easily associated with their true career field through their "core AFSC." While in that career broadening function, their "duty AFSC" is changed to reflect their current position, but their core AFSC remains unchanged. For example, the core AFSC for a weather officer is 15W. If the officer accepts an instructor position, their duty AFSC changes to an 81T designator, but their core AFSC remains unchanged [7]. This makes tracking non-rated officers by their career field relatively simplistic.

Unfortunately, tracking rated officers (specifically pilots, CSOs, and navigators) is not as simple. Rated officers do not necessarily have the same association with a core AFSC as non-rated officers do. They do possess AFSCs, but these are assigned based on the aircraft they are actively flying or maintaining currency in (e.g., tanker pilots are designated 11T while RPA pilots are designated 11U). However, unlike their non-rated counterparts, cross-training into different airframes is quite common in the rated career fields; a tanker pilot could become an RPA pilot and then possess the AFSC, 11U. The Air Force's solution for this career field association issue is a designator known as a Rated Distribution and Training Management (RDTM) code. "Rated officers are uniquely identified by their RDTM code for the purposes of inventory management" [2]. Instead of

a core AFSC, this RDTM code associates the officer with the type of aircraft that they are trained to fly (not necessarily actively flying or maintaining currency in) [2]. The RDTM code helps to address loss of fidelity in career field designation when rated officers cross-train. What the RDTM code fails to overcome, and creates a challenge in retention modeling, is the impact that the culture and lifestyle of a certain airframe has on the retention rates of a specific rated officer throughout the life of their career.

According to a Congressional Budget Office study, “The willingness of pilots to remain in the military depends on many factors... [including the] type of military aircraft flown...” [8]. Rated officers are asked for their preferences for which type of aircraft they will operate, but as with all elements of the Air Force, ranking with respect to peers and needs of the Air Force ultimately dictate the aircraft that a rated officer is initially assigned to. Throughout their career, training into a new airframe is quite common. A tanker pilot who belongs to an airframe with a high retention rate could become an RPA pilot who operates an airframe with an infamously poor retention rate. An officer that may have been willing to remain with the Air Force while assigned to one airframe could become the officer that finds themselves no longer satisfied in a new airframe and more likely to separate. This cross-training occurs with regularity in rated officer career progression, making assignment to AFSC subgroups complicated for the purposes of regression and trend analysis.

In addition to the inherent airframe cross-training issues for tracking rated officers, there is also a concern about the database utilized by the Air Force to maintain personnel records for its members. The data is obtained from personnel record extracts provided by Headquarters Air Force Directorate of Personnel (HAF/A1) from the personnel database

system known as the Military Personnel Delivery System (MilPDS). This system is prone to potential errors in its records due to system glitches, input errors, or missing data. Incomplete or inaccurate data could present a problem in regression analysis, so methods for overcoming these issues are discussed in Chapter 3.

Involuntary separations are the final issue for consideration. These involuntary separations are not only the product of force management programs designed to downsize personnel numbers. Involuntary separations are also the result of poor performance indicators such as failures to maintain physical fitness, disciplinary problems, and medical complications that result in members being forced out of the military. Force management and personnel issues that result in involuntary separation skew the data and thereby make trend analysis more complex or even misleading.

The ultimate intent of this research is to aid HAF/A1 and its supporting agencies in the definition of their manpower policies with respect to rated officers. To ensure that the product created by this research is reproducible within these agencies, the investigation has been limited to the utilization of software packages that are available throughout the Department of Defense (DoD): Statistical Analysis System (SAS), Microsoft Excel software, and Visual Basic for Applications (VBA) code. “SAS is an integrated system of software solutions that enables [a user] to perform data entry, retrieval, and management; report writing and graphics design; statistical and mathematical analysis; business forecasting and decision support; operations research and project management; and applications development” [9]. Excel is a more universally known software package that allows users to organize, manipulate, and calculate data stored in spreadsheets, which can be customized using Microsoft’s event-driven programming language, VBA [10].

1.4 Thesis Outline

This thesis is divided into six chapters. Chapter 2 is a literature review of the past research conducted on employee attrition (civilian and military) with an emphasis on the factors affecting those rates and the models and techniques used to predict future employee retention. Chapter 3 discusses data sources, including MilPDS, the extracts provided by HAF/A1, and sources for economic and political data. Chapter 4 explains the methodology employed and discusses the rated officer retention models. Chapter 5 summarizes the analysis techniques and investigates potential scenarios involving possible shifts in influential economic or political factors. Chapter 6 expounds on the results from Chapter 5 as well as enumerating the limitations to this research and potential follow-on work.

II. Literature Review and Background

2.1 Introduction

Of all the career fields, the rated officer corps (specifically fighter pilots) have seen the largest share of retention studies conducted. This literature review delves into past work by first building off of two antecedent Air Force Institute of Technology (AFIT) theses. From there, the review seeks to explore possible modeling techniques and to identify significant factors to guide a new regression and survival analysis study.

2.2 Antecedent Studies

This work is a follow-on to two recent explorations into officer retention performed by Schofield [11] and Zens [12]. Schofield explored attrition behavior for non-rated Air Force line officer career fields using logistic regression and survival analysis [11]. Zens continued this exploration of the same non-rated officer career fields by applying the results from Schofield's work to predict retention within a current Air Force population and analyze stability of those career fields [12]. This section overviews their work and discusses a potential avenue for expanding the modeling techniques used.

The first study was performed by Schofield [11] who received extracts of officer personnel data from HAF/A1 that spanned the time of January 1999 to December 2013. These extracts captured data for each month within the timespan for a population range between 63,500 to 73,100 officers. Schofield's datasets were filtered to only include non-rated officers within the generic career field classifications of Acquisitions, Logistics, Support, and Non-Rated Operations (these classifications could be further broken down into specific career field subpopulations). These datasets have 315-360 fields per record.

A majority of these fields are either blank or contain extraneous information that prove not significant to attrition rates, e.g., date of birth, social security number, etc.

Retention data yields a binary response variable (i.e., will the officer stay in the Air Force, yes or no?). This binary response led Schofield to utilize logistic regression to identify the significant factors in officer retention within year groups, e.g., 0-6 commissioned years of service (CYOS), 4-8 CYOS, etc. Schofield considered six demographic variables: 1) the year an officer commissioned, 2) gender, 3) commissioning source, 4) number of years served as enlisted personnel prior to commissioning as an officer, 5) specific career field, and 6) distinguished graduate status from commissioning source. After identifying the significant factors, odds ratios were calculated that suggested females (no matter career field or CYOS) were more likely to separate from the military than males. The higher the number of years an officer had as an enlisted service member indicated a higher probability of retention. Additionally, distinguished graduate status was a retention indicator. In contrast, while commissioning source was found to be a significant indicator, officers that were likely to remain in the military varied in commissioning source among CYOS. Career field was also found to be a significant indicator where retention behavior varied by CYOS.

After identifying the significant factors with logistic regression analysis, Schofield used survival analysis to capture retention behavior within each career field with respect to those significant factors. The results for each career field exhibited different indicators for retention and attrition. Overall conclusions from her study suggested that females were more likely to separate, and distinguished graduate status as well as number of years of enlisted service were consistent indicators of retention.

Following Schofield's work, Zens [12] conducted additional analysis on the same career fields that Schofield analyzed. Zens applied the theoretical findings from Schofield's work to subpopulations that were active in the Air Force during the time periods she considered. Zens aggregated the survival function curves for all the career fields as well as for individual career field subpopulations. This aggregation of the survival curves enabled predicting retention rates for the subsequent 30-year period.

Schofield and Zens limited explanatory variables to demographic factors found in the HAF/A1 personnel database. However, research conducted for this follow-on study suggests that retention forecasting should not be strictly limited to demographic data. Discussion of this expanded research will occur in Sections 2.3 and 2.4.

Capon and Chernyshenko [13] conducted a study on retention in the New Zealand military, applying civilian retention theory to a military organization. Previous studies on military retention dismissed the idea of applying civilian retention theory to a military organization due to inherent personnel structural differences in the military. The specific differences are the military's congressionally mandated personnel levels, which bound the number of personnel despite the mission requirements, and the rank structure that does not allow for senior leadership vacancies to be filled outside the organization [11]. Capon and Chernyshenko critique current military personnel forecasting methods for their sole reliance on "demographic (i.e., gender, race, age, and marital status) and organizational characteristics (i.e., male/female ratio, length of overseas assignment)" [13]. The authors point out that while data mining can result in high predictive validities, such an approach does not actually address the root cause of the turnover. For example, "female soldiers

may have a significantly lower retention rate than their male counterparts; however, this does not identify the underlying reason for their behavior” [13].

The current research does not specifically seek to identify root causes for attrition behavior. Rather, the intent is to build a predictive model of rated officer retention of the highest possible fidelity. Capon and Chernyshenko suggest this goal cannot be accomplished with demographics alone and additional factors must be considered. According to Bezdek in Long-Range Forecasting of Manpower Requirements, “forecasting methods... relate requirements for an occupation to different economic, demographic, sociological, and political variables” [14]. Therefore, to obtain a better predictive model, the following literature review will discuss various modeling techniques with an emphasis on the factors that these techniques identify as significant factors in employee attrition.

2.3 Modeling Techniques

Staffing attrition is not solely a military problem. Private and government organizations invest time and financial resources into managing their human capital. Companies and the military both seek to mitigate the loss of these substantial investments and are willing to invest in research to identify the factors that are most likely to be associated with employee attrition. In this exhaustive pursuit to mitigate attrition, numerous modeling techniques have been applied, from mathematical modeling through regression to simulation. This section overviews modeling techniques previously applied to various manpower problems.

An early mathematical model (goal programming) created for simulating and analyzing military manpower was the Accession Supply Costing and Requirements model (ASCAR), originally created by General Electric Company and then revised for use by the

Office of the Assistant Secretary of Defense [15]. ASCAR determined the accession numbers necessary to meet end strengths over a 15-year horizon through a five-step process: 1) analyze historical data, 2) simulate losses based on historical trends, 3) evaluate new recruits based on qualitative and demographic factors, 4) apply goal programming to meet end strength requirements by selection of candidates through qualitative constraints, and 5) assess cost factors for alternative manpower policies. Ideally, this five-step process helps maintain yearly optimum end strength numbers. Unfortunately, goal programming has been criticized by some for not being pareto efficient [16]. “In any multiple objective problem, a solution is said to be pareto inefficient (or dominated) if the achieved level of any one objective can be improved without worsening the achieved level of any other objective” [16]. While there has been some work done to restore pareto efficiency, this lack of pareto efficiency has led researchers to turn to other modeling tools available.

Gass [17] discusses the various types of military manpower models used at the start of the 1990s. The first modeling technique he reviewed was Markov (or transition rate) models. The states were defined by a combination of descriptors of each personnel type (e.g., years of service, job title, skills, specific function), and transition rate models were applied to these states to forecast personnel inventories by estimating new hires, separations, retirements, etc. This Markov modeling technique was limited by its reliance on initial classifications of the personnel inventories and the need for highly accurate transition probabilities that represented rate of state transition. In addition to Markov models, Gass discussed network-flow models and how they were employed to represent personnel systems. This network-flow modeling technique was initially used for personnel scheduling but was later applied to personnel inventories as a minimum-cost network-flow

model that maintained inventory balances from the source node, through the intermediate arcs, all the way to the sink node. “The source nodes [were the] initial inventory, the sink nodes [were the] the final inventory, and the intermediate nodes [were] established to maintain personnel balances and to force inventories to meet grade and skill goals” [17]. Unfortunately, the network model was considered too restrictive when assessing complex and time-dependent analyses.

Because manpower problems are so complex, many argue that these mathematical models are not powerful enough, or require an oversimplification of the problem, to adequately model a manpower system. Additionally, these mathematical models are less flexible to experimentation with the system’s variables and assumptions. Today, many researchers utilize simulation to model real world systems and use the results from those simulations for inference. According to Hill *et al.* [18], the military is already a prolific user of discrete event simulation models “to gain insight into the myriad of issues facing the military... [including] how to structure the military given the uncertainty of the future.”

Simulation is a powerful and flexible tool for manpower analysis and forecasting. In 1999, Collofello *et al.* [19] utilized systems dynamics modeling within the *ithink* simulation software to examine the impact of staff attrition rates on project completion in a civilian software development organization. The authors programmed four basic feedback loops, i.e., “multiple cause-effect relationships connected through a circulation relationship.” These feedback loops were meant to mimic the most influential factors in software development projects. The feedback loops consisted of 1) the staffing profile, such as the number of members on the team and their skill level, 2) the amount of communication that needed to be filtered by each team member, 3) the number of defects

generated that required reworking, and 4) schedule pressure with respect to the completion timelines. This study helped provide insight regarding the impact of attrition and manpower replacement policies on project timelines and budgeting considerations.

Another simulation effort with a more closely related link to this thesis' analysis of rated officers came from Gaupp [20], who created the Pilot Inventory Complex Adaptive System (PICAS) to examine personnel behaviors believed to influence pilot retention. Capturing this behavior was accomplished using a complex adaptive system of interacting pilot agents. PICAS examined both environmental and peer interactions and their influence on pilot retention. The potential preference a pilot would have for staying in or leaving was quantified using three external factors: the availability of pilot jobs in the commercial airline market, the perceived pay gap between military and commercial pilots, and the current flying operations tempo in the military. As Gaupp pointed out, "the behavior of the actual pilot-agents [in the simulation] is controlled by means of shifting a set of two utility curves, one representing the amount of satisfaction that the pilot-agent receives from money, the other representing the amount of gratification the pilot-agent receives from time-off" [20]. Examining the combination of these factors revealed that time and operation's tempo (ops tempo) were the significant factors in pilot retention rates. However, his study was not intended to actually predict retention so much as capture the potential effects of external factors on pilot attrition behavior. These effects identified by Gaupp regarding external (e.g., economic) impacts on the propensity for rated officers to retain are worthy of further examination and modeling.

While simulation is a great tool for mimicking complex constructs of real world systems and aiding analysts in making inferences regarding the behavior of those systems,

it still requires the analyst to use assumed probability distributions on various system components. The simulation is less useful without an analysis of data and historical trends to create accurate probability distributions for observed events, some of which are binary in structure. For these reasons, manpower analysts often turn to the modeling techniques of logistic regression and survival analysis. Logistic regression is more appropriate than ordinary least squares regression when the response variable is binary because logistic regression “does not assume normality of the data, linearity of the relationship between the independent and dependent variables, or homoscedasticity” [21].

An example of logistic regression analysis for personnel modeling is Hall [21]. Hall sought to improve retention rate modeling for the Army Dental Corps, which was seeing a higher than expected attrition rate from its dentists at the end of their service commitment [21]. Hall focused on the impact on Army dental officer retention when the dentists were confronted with an increased ops tempo confounded by an attractive civilian healthcare job market. The data sets used consisted of information on all Army dentists on active duty from 1998 to 2008. The data sets were used to create a useful model for predicting Army dentist retention. In addition to examining the effect of ops tempo on attrition, Hall included age, sex, race, family status, accession source, and residency completion. Conducting a properly executed logistic regression analysis can be time-consuming and tedious as the final mathematical model is found by including and excluding each potential explanatory variable and then assessing the goodness-of-fit. Hall assessed seven potential models before selecting the most parsimonious model with the best predictive capability. His results suggested that age, race, and family status were statistically significant, but gender and ops tempo were not [21].

As evidenced by several of the previous research examples, data sets used for the studies are usually excerpts of large datasets spanning between ten and twenty years. However, personnel inventories are an ever-changing system with new accessions and separations occurring continuously. Due to this dynamic property, any excerpt of time from a personnel database will contain incomplete records. This means that some personnel records are either censored, meaning that the “observation’s full event history is not observed,” or truncated, meaning “a period of an observation’s history is not observed” [21]. Unfortunately, logistic regression typically excludes observations that are outside of the analysis time due to either censoring or truncation [21]. Therefore, many manpower analysts incorporate survival analysis, which can accommodate censored data [11].

Hall [22] studied the attrition rates of the Marine Corps’ enlisted personnel using survival analysis. He assumed that the longer a service member remained on active duty, the less likely the service member was to separate. His data sets included accession data, personal statistics, and separation data for all Marines who enlisted between 1996 to 2008. Previous studies showed that attrition rates for Marines decreased after 12 years of service, so he focused his study on attrition behavior from initial year to 12th year of active duty, which consequently contained censored or truncated data. Hall created a model using gender, rank, citizenship, race, and educational level as significant factors within each occupation field. These factors did not have consistent effects within each occupation field, so he determined that application to subpopulations (i.e., occupation field) was more appropriate and was able to link variable factors to attrition rates within those subpopulations [22].

2.4 Significant Factors

Bezdek [14] identified four categories of variables as key variables for manpower forecasting: economic, demographic, sociological, and political variables. This section specifically discusses factors from each of the four categories that were identified by other research as significant to attrition rates, starting with economic variables.

According to the Bureau of Labor Statistics (BLS) Military Job Outlook, “When the economy is thriving and civilian employment opportunities are generally more favorable, it is more difficult for the military to meet its recruitment quotas... During economic downturns, candidates for military service may face competition” [23]. The connection between the economy and retention is no revelation. The Air Force is well aware of this crucial factor and even published a very similar message to the BLS’s outlook in its own AFIs: “The economy, particularly the aviation industry for pilots, has historically been the primary determinant of aircrew retention, and the AF has limited influence on this external factor. If the economy is strong, AF retention usually suffers, and vice versa” [2]. In fact, the Air Force has thoroughly identified some of the key factors to rated officer attrition using exit surveys. Those factors are the economy, a high ops and personnel tempo (political), turbulence (i.e., “unpredictable and short-notice taskings,” which are likely due to political circumstances), and erosion of pay and benefits (economic, political) [2].

Pilot retention has been an issue for the Air Force during economic downturns (e.g., 2004-present). A recent study performed by Sweeney [24] provides projections that major airline hires will steadily increase until 2026 while the number of available pilots in the civilian sector will not be sufficient to meet the demand. Additionally, the current unemployment rate is at its lowest rate since 2007 at 4.9%, with some economists saying

that we have reached “full employment,” which will lead to a demand for employees that will spur wage increases [25, 26]. Despite being offered very generous aviation continuation pay (ACP) bonuses to stay (some fighter pilots were offered one-time retention bonuses as large as \$225,000), the take-rate for these bonuses has been on the decline since 2010 [24, 27]. In an article in *AIR FORCE* magazine, pilots cited their top 5 reasons for declining these bonuses as 1) high ops tempo, 2) quality of life factors, 3) desire to fly for airlines, 4) commitment too long, and 5) assignment process [28].

A study conducted by Jantscher [29] identified economic factors that were potentially influential to officer retention. She utilized personnel data from 100 different AFCSs over the time period of 2010 to 2014 in combination with economic sites such as the Bureau of Economic Analysis (BEA) and Quandl Financial and Economic Data (Quandl). From the BEA, she extracted quarter and annual information regarding Gross Domestic Product (GDP), and from Quandl, she considered data in a broad range of categories such as growth, employment, inflation, etc. While some AFSCs were linked to additional factors, the significant economic indicators that were seen in virtually all AFSCs were unemployment rate and youth unemployment rate; only two AFSCs (Chaplain and Intel) defied the expected trend of high attrition rates when the economy was strong [29].

In addition to the economic factors, demographic factors are potentially important. Examples of demographic factors include race, gender, marital status, dependent status, career field, etc. One factor that has consistently found to be a significant factor is gender. An issue paper from the Military Leadership Diversity Commission (MLDC) reaffirmed this conclusion about female attrition rates when they stated, “Continuation rates among women are lower than among men. On average, regardless of Service branch, women’s

CCRs [career continuation rate] are lower than men's for every YOS beyond YOS 2 or YOS 3" [30].

Another consistent demographic factor is marital status. According to Cerman and Kaya [31], who studied the effect of marital and family status on retention and promotion of Marine Corps officers, "married male officers obtain higher fitness report scores, higher promotion probabilities, and higher retention probabilities than single officers" [31]. Furthermore, they found that, "having additional non-spousal dependents increase fitness report scores and retention probabilities."

The significance of race, however, has been substantiated less in research results. While two studies [21, 22] concluded that race was a significant factor, the study from the MLDC [30] determined that with respect to race, "on average, blacks' and Hispanics' CCRs are greater than or equal to whites' rates at every YOS point, while Asian/[Pacific Islanders'] and others' rates are less than or equal to whites' rates. There are some exceptions (most notably, blacks in the Air Force), but in general, these data indicate that retention among minority officers is not lower than among whites, which indicates that there is no immediate need for a policy response" [30]. Salas [32] came to the same conclusions with regard to Hispanic officers. Specifically, Salas determined that, "Hispanic Marine Corps officers have a greater likelihood of retention but no difference in fitness report performance and no difference in the probability of promotion to O-4 in comparison to non-Hispanic officers" [32].

Within the purview of sociological factors, two studies specifically addressed the person-organization fit. Carter [33] explored the relationship between person-organization fit and attrition patterns for active duty Army officers with less than eight years of active

duty service. For her data, Carter collected various research sources with similar topics, sorted and ranked the studies with respect to relevance to her investigation, and aggregated the data from the relevant studies to create a master dataset. The subsequent meta-analysis investigated qualities such as “fit, job satisfaction, organization commitment, and intent to stay... [as well as] attitude toward career field, being able to use one’s abilities, job challenge, job involvement, job excitement, job autonomy, job enjoyment, family support, career mobility, tenure, values, level of camaraderie, workload, feedback, preparation for future responsibility, leadership quality, retirement benefits, and organization level” [33]. Among these qualities, Carter found that fit, job satisfaction, organization commitment, and intent to stay were significant factors in predicting officer retention.

Caswell [34] conducted a Delphi study using 20 active duty female pilots. Caswell had the pilots answer three sets of questions (five to eight questions each) that elicited the factors most likely to motivate these officers to remain or separate from the Air Force. Just as Capon and Chernyshenko [13] suggested in their application of civilian theory to military organizations, this approach is more likely to get at the root cause of officer retention (and specifically female officer retention). Caswell learned that a lack of career spouse support in the form of joint spouse programs concerns, stress to family integrity, and work/family balance were the largest motivators of female attrition [34].

While in both [33] and [34] sociological factors are extracted that may more likely point to the root cause of attrition rates, the extrapolation of this data and implementation of these methods are not practical. As Hall [22] pointed out, “an organization cannot [identify these root causes] in any practical fashion. Unless all employees could either be continually surveyed for job satisfaction or managers become mind readers, organizations

cannot identify who would leave. Continual surveys are inefficient and mind reading is impossible.” For this reason, sociological factors are not addressed in this study’s pursuit of a predictive model. The data are not available in any consistent, quantifiable format, and the practice would not be practical to implement.

2.5 Logistic Regression

According to Hosmer *et al.* [35], ordinary least squares regression is the primary method for estimating unknown parameters because the resulting estimators generally possess desirable statistical attributes. Unfortunately, if the regression is being applied to data with a binary response variable, then these desirable attributes are lost. In other words, logistic regression is preferred over ordinary least squares regression because it “does not assume normality of the data, linearity of the relationship between the independent and dependent variables, or homoscedasticity” [21]. In retention studies, the response variable is a binary dataset that answers the question, “Will the officer retain, yes or no?”

As mentioned when discussing Hall’s [21] logistic regression work, conducting a properly executed logistic regression analysis can be time-consuming. It requires an iterative process of including and excluding each potential explanatory variable and then assessing the contribution to the log-likelihood function. The contributions that maximize the likelihood function are incorporated into the model; however, including one variable could mean that the likelihood function is maximized by excluding another that had already been incorporated. This process could be quite tedious and time-consuming if an analyst has many explanatory variables that they wished to investigate. Fortunately, the statistical software suite, SAS, expedites this iterative process with the logistic regression function *proc logistic*.

While logistic regression modeling can be used for manpower studies, it is typically considered not as informative as survival analysis because it excludes observations that are outside of the analysis observation period due to either censoring or truncation [21]. This study utilizes the logit model in the limited capacity of assessing the significance of the demographic variables and the fit of the model. To that end, several terms regarding model fit are discussed.

First, “pairs” are defined as “the total number of distinct pairs in which one case has an observed outcome different from the other member of the pair” [36]. These pairs are then assessed to see “if the observation with the lower ordered response value... has a lower predicted mean score than the observation with the higher response value” [36]. If so, then the pairs are labeled “concordant.” If not, then they are “discordant.” If the estimated probabilities are identical, then they are considered “tied.” The percentage of concordance within a model is generally an indication of model quality. If the percentage of concordance is 0.5, then that is considered a poor model with a random probability of accurately estimating the response variable. The closer the percentage of concordance gets to the value 1.0, the better the model is at estimating the response value.

Another concept related to model quality is the c-value. This value is equivalent to the Receiver Operating Characteristic (ROC) curve. “The area under the ROC curve... provides a measure of the model’s ability to discriminate between those subjects who experience the outcome of interest versus those who do not” [35]. Both the c-value and the area under the ROC curve range from 0.5 to 1.0. The closer to 1.0, the better the model performs at predicting the response value accurately.

Finally, odds ratios are a particularly informative aspect of logistic regression that aid in interpreting the effect of a covariate. “The odds ratio is widely used as a measure of association as it approximates how much more likely or unlikely (in terms of odds) it is for the outcome to be present among a subject [with covariates equal to one value compared to the same covariate at another value]” [35]. For example, the research conducted in [30] found that females in the military retain at a lower rate than males. If the odds ratio of retention for females was 0.5, then the odds of females retaining would be one-half that of males (ignoring all other factors). This could be particularly useful in measuring the relative risk of a subgroup when exploring retention data.

2.6 Sustainment Line

While the Defense Officer Personnel Management Act (DOPMA), which is overseen by Congress, dictates the overall number of officers authorized in the military each year, the accession and retention of these officers is regulated by HAF/A1’s Personnel Division [37]. To manage retention and meet Congress mandated levels, HAF/A1 [38] models retention patterns by calculating the percentage of personnel retained at the end of each year. These retention percentages are calculated for each AFSC’s year group, and the last five years are averaged together to give a simple average retention rate for each year group in that career field. These averaged retention rates are also aggregated together to give an overall retention rate for the Air Force. Figure 1 shows an example of a career field health chart extracted from the June 2016 HAF/A1 career field health briefing.

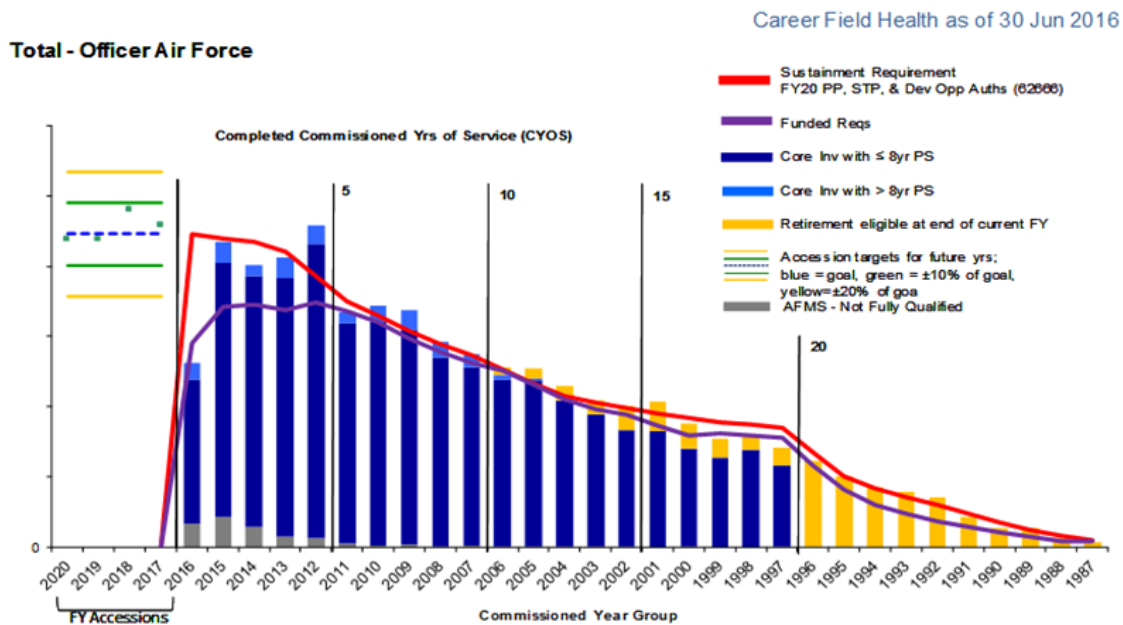


Figure 1. Career Field Health Chart (All Officers)

In Figure 1, the sustainment line is depicted by the red line, which is a mathematically “smoothed” version of the sustainment line calculations previously explained. The purple line just below it is the funded requirements line (i.e., the number of officers that the career field requires in each year group in order to meet current mission requirements). Ideally, the red line should exceed the purple line to ensure that mission requirements are met. When that is not the case, the career field is considered undermanned or “stressed.” The bar graph depicts current manpower inventories based on commissioned year group. Finally, the yellow and green lines to the left of the bar graph represent the desired number of accessions for the upcoming fiscal years. During the course of an officer’s career, fellow officers will attrit from the service due to personal desires, unsatisfactory performance, or mandatory reductions of manpower due to DOPMA and/or mission needs. Maintaining an adequate accession goal ensures that the Air Force has a

large enough pool of young officers to produce a desired number of Field Grade (Major to Colonel) and General Officers.

All of these factors are managed by HAF/A1 to ensure that current accession goals ensure current mission needs are met. Unfortunately, this model does not account for potential changes in attrition rates due to internal and external (e.g., economic) factors. This model strictly represents the historical retention rates of its personnel assuming that mission needs and external factors do not change. If HAF/A1 is aware of a change in the economy, it can guess to the impact that such a shift might have to retention rates and tweak the model according to its guess; however, the model itself has no built-in capability to capture the impact of economic factors on retention. To overcome this deficiency in modeling flexibility, survival analysis is proposed as an alternative to current HAF/A1 methodology.

2.7 Survival Analysis

Unlike the methodology described in Section 2.6, survival analysis incorporates explanatory variables (also known as covariates). The random variable in survival analysis represents time, or specifically, the time to an event. In this study, the random variable of interest is the time that elapses until a rated officer separates from the military (i.e., the retention rate). While parametric methods of applying a known probability distribution to fit the behavior of a random variable can be very informative, “nonparametric methods provide simple and quick looks at the survival experience” [39]. The survival function is a simple transformation of any dataset’s cumulative distribution function (CDF). The CDF describes the probability that an event will occur before or at a given time while the survival function captures the probability that the event will occur after a given time. This survival

rate of an officer's career is of interest because the Air Force wants to know how long they can expect to retain an officer. In other words, what is the minimum amount of time an officer can probabilistically be expected to retain?

The initial steps of survival analysis typically use non-parametric methods such as Kaplan-Meier (KM) estimators to capture the survival function without aid of a known probability distribution to fit the data. This mirrors the technique utilized by HAF/A1, as it simply correlates the probability of survival to the amount of time that has elapsed. Subset populations can be compared by generating KM curves for each subset on one graph. Figure 2 [39] is a graph generated by SAS with two KM curves from two separate subpopulations overlaid on one graph.

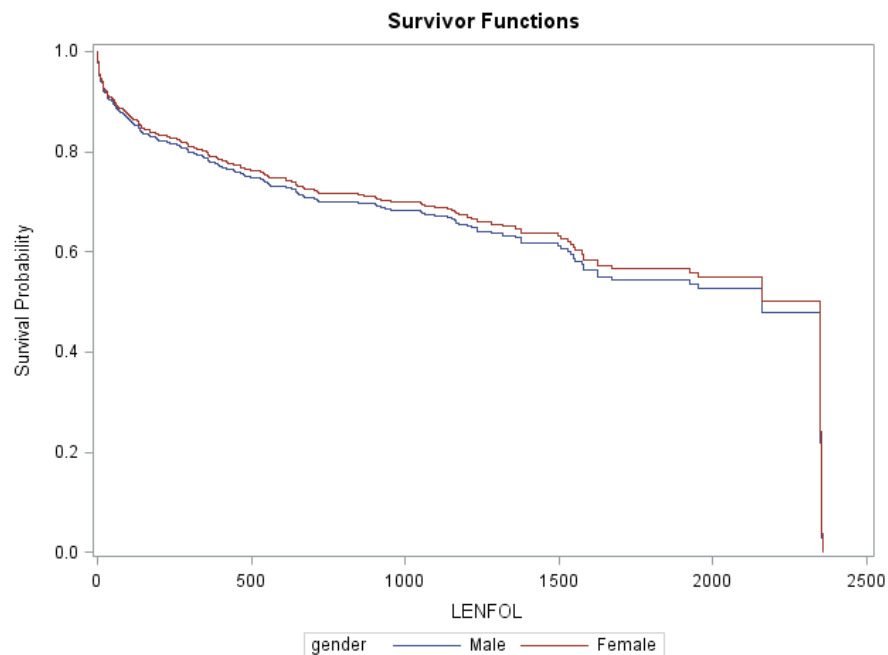


Figure 2. Example KM Curve [39]

In Figure 2, the survival rate starts at 100% at time = 0 in the top left corner. As time progresses along the x-axis, the survival percentage decreases until it reaches the end of the observation period at 2500 days. Comparison of the two lines shows that the survival

rate for males begins to decline at a larger rate than females starting at approximately the 400-day mark. While they have similar curve shapes, the percentage of survival for males (if significant) is lower throughout the observation period compared to females; however, no correlation to external factors can be gleaned from this model.

To incorporate explanatory variables, Cox Proportional Hazards Regression (PHREG) is employed. This is now an examination of the hazards function, which “gives the probability that the subject will fail in that interval, given that the subject has not failed up to that point in time” [39]. Cox PHREG removes the dependence on time by expressing the hazard rate as a product of two functions. One function still describes the relationship between the hazard rate and time while the other describes the relationship between the hazard rate and the covariates.

There are two methods in SAS for determining significant covariates. First, if covariates are inputted into the model indiscriminately, then SAS assesses the parameter estimates for fit, and the user applies a significance level test. In this study, the significance level is 0.05. The second method is to employ the stepwise option. The stepwise option starts with no covariates in the model. It adds covariates one at a time, assesses the model fit with the covariate added, and then either incorporates or rejects that covariate. It performs this process iteratively until it has added all covariates to the model that fall within user-specified, significance level tolerances. Either of these models that incorporate covariates into the hazard function can then be manipulated to see the effects of certain parameter values.

III. Data Sources

3.1 Introduction

As found in the literature review, demographic, economic, and political variables may be useful for improved manpower forecasting [14]. This chapter reviews the strengths and weaknesses of the data sources for those variable types.

3.2 Demographic Data

The primary resource for the demographic data in this study is the Military Personnel Data System (MilPDS). According to Diversified Technical Services, Inc., which provides development and sustainment support for the MilPDS, “[The system] is an Air Force wide, enterprise-scale military system constituting one of the world’s largest Oracle Human Resource (HR) implementations, involving over 100 military subsystems and in excess of 500,000 business rules, supporting all active duty and retired USAF members as well as Guard and Reserve components” [40]. MilPDS is a web-based application associated with headquarters at Air Force Personnel Center (AFPC) at Randolph Air Force Base, and it has a global customer base. Essentially, where there is an internet connection and a properly trained personnelist, a service member can access this system and update/maintain their personnel records.

More specifically, the Military Personnel Records System AFI [41] defines MilPDS as the central database for all military personnel records data, and this database can be manipulated through actions initiated at the user-, technician-, or headquarters-level. MilPDS contains all information pertaining to a service member’s career such as name, social security number, date of birth, rank, date of rank, assignment history, marital status,

dependent status, education information, decorations, performance reviews, disciplinary actions, etc. There are over 300 data fields for each service member, and this data information is vital in assisting the commanders as well as the member in the maintenance of their career.

Unfortunately, this database is subject to occasional errors. Changes to this system are made constantly as technicians across the globe input new information and update current records. Automatic updates occur on a frequent basis for routine actions, e.g., promotion to the next rank on the anniversary of a date of rank. Additionally, necessary and vital upgrades to the system could result in unexpected system malfunctions that could alter or delete data fields. This ever-changing database is vulnerable to input or deletion errors in one or many records at any given time. In an effort to combat data loss, the Air Force distributes notifications during these upgrades with guidance that requests members to review and manage their records to ensure data alteration or loss is corrected as soon as possible [42]. Additionally, data back-ups, referred to as “snapshots” are captured on a frequent basis and are archived for reference in case of minor or wide-spread system issues [11].

The data used in this study consists of personnel records for the active duty rated officers in the AFSCs of Pilot, CSO/Nav, and ABM from January 2006 to December 2015. These monthly snapshots of rated personnel records were extracted by HAF/A1 from MilPDS and converted to SAS format for ease of use. Since HAF/A1 is aware of the potential for incomplete or erroneous data fields in the extracted data, it developed programs that automatically fix some of those errors and also convert rated officer’s RDTM codes into rudimentary core AFSCs [11]. These corrections help to alleviate data

inconsistencies that may be encountered during regression analysis. Unfortunately, as detailed in Chapter 4, corrective actions do not address issues with data truncation or censoring, which are common when using extracts of personnel records. Selection of an adaptable modeling technique and proper refinement of the data are the best aids in overcoming these censoring issues.

3.3 Economic Data

Private and government institutions acknowledge that the strength of the economy influences their ability to obtain and retain employees [2, 14]. When the economy is strong, businesses are healthy, seeing larger profits, and typically looking to expand and hire more workers. When the economy is weak, businesses typically shrink and reduce their workforce. For the military, this means “when the economy is thriving and civilian employment opportunities are generally more favorable, it is more difficult to meet its recruitment [and retention]” goals [23]. Therefore, being able to predict the strength of the economy at any given time can be an insightful method of predicting retention patterns. Unfortunately, forecasting the troughs and peaks of an economic cycle is a complex process that macroeconomists are still struggling to model [43].

For economists, economic variables are distinguished by their direction and timing. With respect to timing, there are three types of economic variables: lagging, coincident, and leading indicators [29]. Lagging indicators are measurable effects that occur after economic activity. Examples of lagged indicators include gross domestic product (GDP), income and wages, consumer price index (CPI), currency strength, interest rates, and even corporate profits. Because these are indicators that occur after the advent of the economic

phenomenon, these factors are not useful for prediction of trends, but they could confirm the existence of long-term trends.

Leading and coincident indicators, however, are quantifiable statistics that change at or before the economy starts to follow a particular pattern [43]. Unlike lagging indicators, these indicators could be used to predict future economic trends, but they are only useful if they are accurate. Examples of leading indicators include the Durable Goods Manufacturers' Shipments, Inventories, and Orders Report, the Purchasing Managers Index ® (PMI), and the Consumer Confidence Index ® (CCI). What separates these reports from the lagging indicators above are their investigation into U.S. consumers' expected future expenditures. According to the BLS, "personal consumption expenditures... account for more than 60 percent of total employment in [the] U.S. economy" [44]. Fluctuations in these indices could be strong indicators of future economic strength; however, the science behind some of these indices is questionable.

The Durable Goods Report (DGR) (current and historical data) can be obtained from the U.S. Census Bureau [45]. This report is released once a month and includes a one-page summary of the significant changes in new orders, shipments, unfilled orders, and inventories for over 4,000 manufacturers of durable goods. Data taken from the DGR are leading variables with concrete values. Alternatively, the two example indices from above come from private institutions. The Institute for Supply Management publishes the PMI® once a month and bases its index on five major indicators: new orders, inventory levels, production, supplier deliveries, and employment environment [46]. The CCI® is a metric created by a business research association known as The Conference Board. A monthly, randomly sampled survey is conducted by the analytics company, Nielsen, to

capture consumers' impressions of current economic conditions [47, 48]. These two indices, the PMI® and CCI®, illustrate the potential issues with relying on leading indicator indices: they are constructed with fuzzy science, relying on the responses obtained from surveys from consumers.

Finally, according to Abel and Bernanke, “although the Conference Board [an independent global business research organization] has studied the timing of unemployment... the timing of this variable is designated as ‘unclassified,’ owing to the absence of a clear pattern in the data” [43]. It neither assumes leading or lagging characteristics. Jantscher’s [29] research noted that many researchers consider the unemployment rate to be a leading economic indicator. Jantscher’s research confirmed the results of other studies and determined that unemployment rate was a significant indicator of Air Force retention.

3.4 Political Data

Political data may be the most difficult to quantify and integrate into a mathematical model. The main overarching theme for this section is the militaristic opinions of the electorate and how that is reflected in its governing body (i.e., the policy makeup of the presidential and legislative branches). An electorate that preferred fewer warfronts and reduced enemy engagement for the sake of its soldiers could result in a lower ops tempo. In contrast, an electorate that is more willing to act offensively rather than defensively could result in more warfronts and a higher ops tempo.

While the reflection of the public’s isolationist and engagement mentality may be inferred from the makeup of the executive or legislative branch, the personnel data extracts obtained for this study constrain the sample populations for these governing bodies.

Because the president's tenure can be up to eight years, and the data extracts for this research spans only two different presidential administrations, it is doubtful any valuable inferences could be made with respect to who occupies the position of president. Furthermore, even with mid-term elections ensuring a higher turnover rate and a larger population in the U.S. Congress, this change in the legislative body only occurs every two years. Changes every two years within a ten-year span means that there are only five unique legislative combinations within the timeframe being examined. Again, this sample population is likely too small to make any valuable inferences.

The best option for examining political factors may be through ops tempo, perstempo, and male-to-female ratio for officers. Ops tempo is the operational activity of soldiers while at their home station whereas perstempo is the number of "deployment days an individual spends away from home station... or their assigned unit" [49]. Ops tempo varies from base to base due to mission and airframe training requirements. Because of the specific nature and lack of standardized records between units for this type of data, this is not an ideal statistic for regression or survival analysis. Future research could perform site-specific studies of ops tempo to determine impact on retention. However, a representation of perstempo can be extracted from MilPDS through the duty status designator. According to the Duty Status Program (AFI 36-2134) [50], personnel that are deployed on contingency exercise deployment (CED) orders are assigned the duty status code '20.' The percentage of officers deployed monthly could be used to determine a perstempo rate to compare against attrition rates. Finally, the exact male-to-female ratio from any given monthly extract can also be calculated for comparison with attrition rates.

IV. Models of Rated Officer Retention

4.1 Introduction

Logistic regression is first used to identify significant predictive factors and analyze potential retention patterns. Survival analysis is then used to create a model for predicting retention.

4.2 Logistic Regression

4.2.1. Introduction

Binary logistic regression, or logit regression, is employed to determine whether certain demographic variables are statistically significant in predicting the retention of rated officers. Odds ratios for those significant variables are extracted.

4.2.2. Data

The data used for the logistic regression is an aggregate of all monthly extracts provided by HAF/A1 in SAS file format for the period of January 2006 to December 2015. The extracts are provided in two forms: current inventory and losses for each month. A “retain” variable is added to both the inventory and the loss data. For overall assessment of each career field, the retain variable is assigned a ‘0’ in all loss files and a ‘1’ in all inventory files. This retain variable is the binary response variable for the logit model. To combine the years, the last month of each year and all of the year’s monthly losses are appended together.

The data are refined after all files are aggregated using basic SAS commands (see Appendix A for code). The extracts contain all officers in the Air Force, so the files are reduced to rated officers only (identified through the RDTM codes). Within the rated

officer data are student rated officers. These are the officers that are currently going through their respective training programs (e.g., Undergraduate Pilot Training) and have not qualified as rated officers yet. The retention of these student officers is not subject to the same factors as those that have completed their training. Rather, the retention of these officers is based upon individual motivation and performance in the competitive training environment. This study does not seek to identify retention patterns among rated officers in the training environment. Therefore, student rated officers were excluded from the subsequent regression analysis.

Further refinement of the data includes deletion of duplicate records and redefining explanatory variables. For duplicate records, the last record is retained. It is assumed that the last record (last recorded month in the officer's career) is the most complete and accurate. From there, the demographic data most likely to be significant based upon the literature review is redefined for ease of use in the regression. The covariates that were included in the model were Marital Status ('0' = Single, '1' = Married, '2' = Divorced/Separated/ Widowed/etc.), Gender (Male or Female), Source of Commission ('1' = Other, '2' = U.S. Air Force Academy (USAFA), '3' = Reserve Officer Training Corp (ROTC), '4' = Air Force Officer Training School), Distinguished Graduate at source of commission ('0' = Regular Graduate, '1' = DG), Prior Enlisted Service ('0' = Did Not Have Prior Service Time, '1' = Had Prior Service Time), and whether the officer had dependents ('0' = No Dependents, '1' = one or more Dependents).

Finally, the data are separated into career field subpopulations (e.g., pilot, CSO, ABM) and further sorted into subsets of each career field based upon commissioned years of service (CYOS). The range of CYOS overlapping subsets includes 0-6 CYOS, 4-8

CYOS, 8 -14 CYOS, 12-19 CYOS, and 20-22 CYOS. Unfortunately, these personnel records are inherently prone to censoring (“response values cannot be observed for some or all of the units under study”) and truncation (“observations are actually observed only when they take on values in a particular range”) [51]. Because these extracts are just a snapshot in time for personnel records, some officer’s records are not complete. This is because the beginning, the end, or both the beginning and end of a military career may not be encompassed in the ten-year span of the records being examined. Reducing both censoring and truncation would severely diminish the richness of the data, so efforts are taken to reduce the truncation of data for the analysis. To ensure that data are not truncated, officer’s records are assigned to each CYOS subset only if that record spans the entirety of the CYOS subset’s range. For example, if an officer ends their career after 17 CYOS, then that officer’s records will belong to the 0-6 CYOS, 4-8 CYOS, and 8-14 CYOS subsets. However, observations are not observed for all values in the range of 12-19 CYOS, so that officer’s record is not included in the logistic regression for 12-19 CYOS.

4.2.3. Analysis at Career Field Level

Once the refinement of the data is complete, SAS’s *proc logistic* command is utilized to determine a model for rated officer retention. All covariates used in the model (listed in Section 4.2.2.) are class variables (either categorical or binary).

First, analysis of the data at the career field level (e.g., pilot, CSO, ABM) is conducted. Based upon the literature review, we were not expecting strong correlation or model convergence; however, all three of the career fields did actually produce logit models that converged under SAS’s relative gradient convergence criterion. Unfortunately, while the ABM career field did converge at the career field level, it

produced some of the weakest correlations once broken out into CYOS subsets. Additionally, all three models produced stronger model fitness scores with the covariates included in the model as opposed to the model with only the intercept. Table 1 summarizes the Wald Chi-Square p-values for the covariates for each of the career fields (significance level of 0.05).

Table 1. Logistic Regression Covariate p-Values for Pilot, CSO, and ABM

Covariates	Pilot	CSO	ABM
Marital Status	<0.0001	<0.0001	0.0023
Gender	<0.0001	<0.0001	0.0001
Commissioning Source	0.0021	<0.0001	<0.0001
Prior Service	<0.0001	<0.0001	<0.0001
DG	<0.0001	<0.0001	<0.0001
Dependents	<0.0001	<0.0001	0.0001
Percent Concordant	60.6	66.0	70.4
Percent Discordant	24.2	20.2	20.1
Percent Tied	15.2	13.8	9.5
c	0.682	0.729	0.751
Number of Observations	11,752	4,090	1,162

This initial logistic regression analysis indicates that all variables included in the model are significant indicators for retention within all three career fields. This analysis includes the concordant and discordant percentages and c-value with respect to the predicted probabilities and observed responses. The c-values shown in Table 1 indicate that the model's predictive capability is better than random chance with values ranging from 0.68 to 0.75. A value closer to 1.0 would have been ideal, but considering the potential noise and censoring issues of using this real-world data extract, these are satisfactory results. This assessment is reinforced by the combination of concordant and

tied percentages, which are nearly 80%, and discordant percentages, which are all in the low 20% range.

In addition to the p-values, the odds ratio of retention for each of the six covariates are examined with respect to the three officer career fields and their likelihood to retain over a 20-year career. For example, Figure 3 details the odds ratios of retention for the marital status variable. Single officers are used as the baseline of comparison and their values are set to one. Among the three career fields, married officers and previously married officers are more likely to retain compared to single officers. In fact, the odds of retention are double or even triple that of single officers in all career fields.

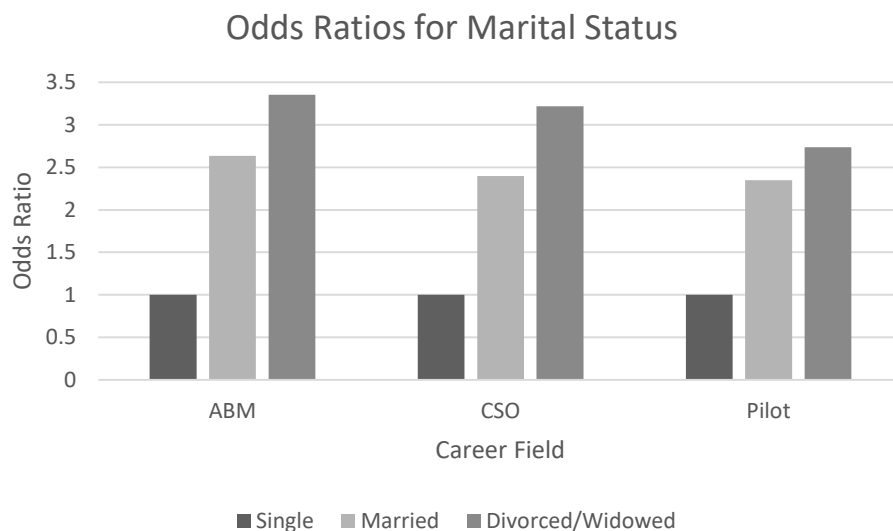


Figure 3. Odds Ratio of Retention by Marital Status

Figure 4 gives the odds ratio of retention with respect to gender within the three career fields. Again, the results for all three career fields confirm that female officers are less likely to retain for 20 years compared to their male counterparts. Female CSOs have

the lowest odds ratio at 0.157 while female ABMs are closest males with a ratio of 0.451.

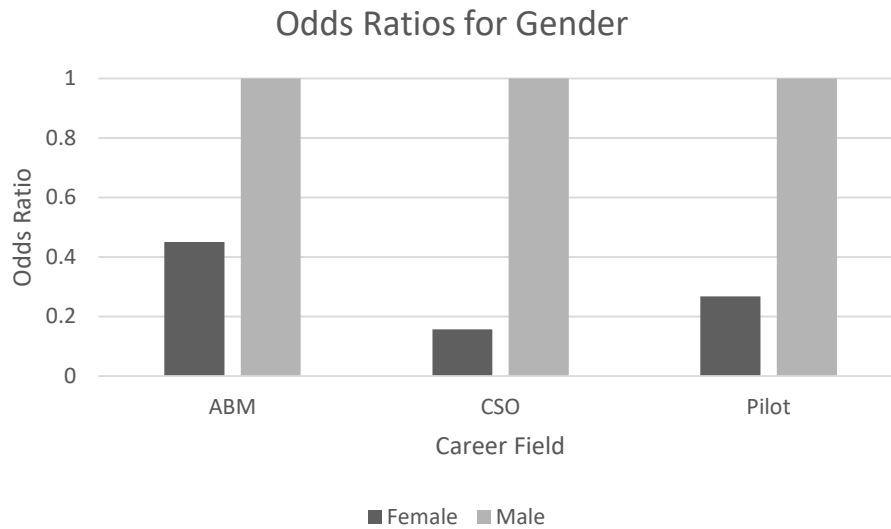


Figure 4. Odds Ratio of Retention by Gender

The odds ratio of retention based on commissioning source is shown in Figure 5. The baseline for the commissioning source is OTS (kept at a value of one). Commissioning source is the only variable whose odds ratios have confidence intervals that cross 1, which means that they are not statistically significant. For all three career fields, the odds ratios for other commissioning sources are not statistically significant. For ABM and CSO, ROTC odds ratios are not significant; for pilots, USAFA odds ratios are not significant. For those results that are significant, Figure 5 is still informative. The odds ratios indicate that USAFA ABM and CSO officers are consistently more likely to retain for an entire 20-year career compared to the OTS graduates (ABM is highest with an odds ratio of 3.50). For ROTC pilots, retention is slightly less likely than for OTS pilots with an odds ratio of retention of 0.858.

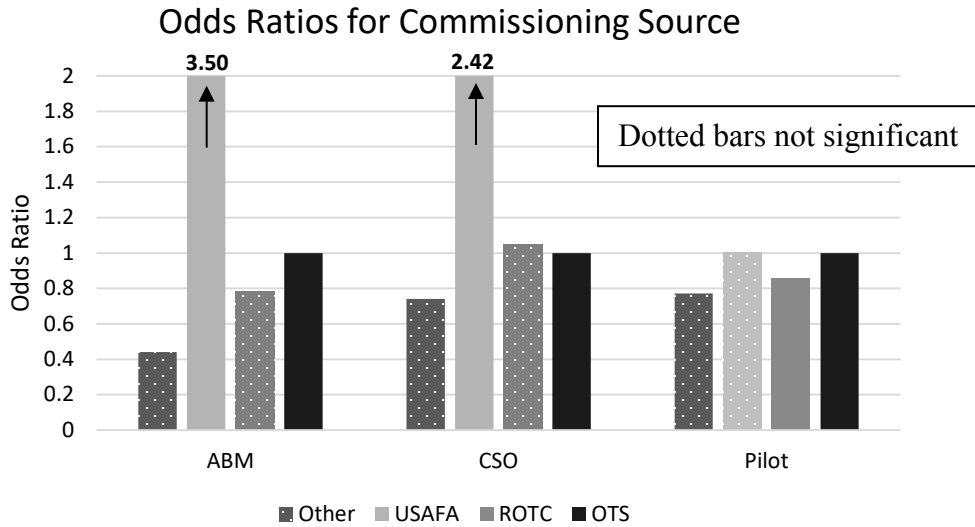


Figure 5. Odds Ratio of Retention by Commissioning Source

Figure 6 shows the odds ratio of retention for prior service officers within each of the career fields. This portion of the analysis carries a key assumption; analysis of officer retention focuses on the probability of completing 20 CYOS not 20 years of total active federal military service (TAFMS). Focusing on CYOS retention is not necessary for nonrated officers because it is easier to train and then inject an officer into a new career field. Because of the intense training and experience requirements of these rated jobs, officer retention is most cost-efficient if the entirety of a 20-year rated and commissioned career is served [4]. Whenever a prior enlisted member becomes a rated officer, that previous enlisted time accumulates in their TAFMS. An officer may retire at 20 years of TAFMS. This means that retention of a rated officer for 20 CYOS is less probable if the officer has prior military service because they will accumulate 20 TAFMS years before 20 CYOS. This produces the results seen in Figure 6. For comparison, Figure 7 shows the odds ratio for prior enlisted officers with respect to a 20-year TAFMS as opposed to Figure

6 with 20 CYOS retention. Clearly, prior service officers in all three career fields are more likely to retain for 20 years of service, but the USAF does not necessarily recoup their investment with a 20-year rated career from these prior service officers.

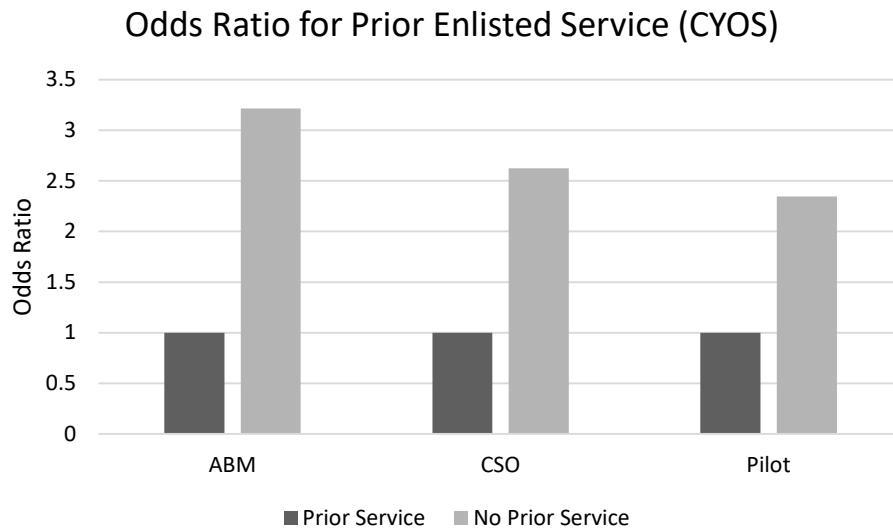


Figure 6. Odds Ratio of Retention by Prior Enlisted Service (CYOS)

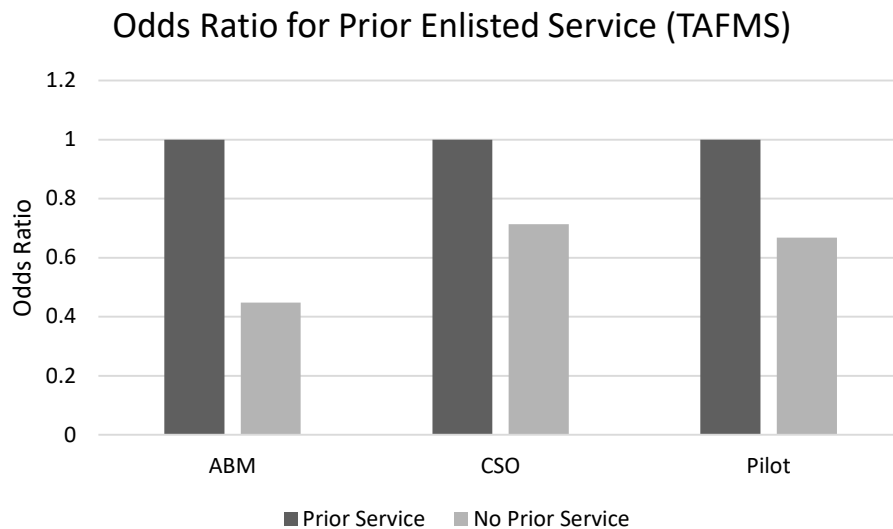


Figure 7. Odds Ratio of Retention for Prior Enlisted Service (TAFMS)

Figure 8 shows the odds ratio of retention for officers who earned DG at their commissioning source versus regular graduates from commissioning sources. The baseline for this odds ratio is officers who were DGs at their commissioning source. Among the three career fields, likelihood of retention for a 20-year career is closest between DGs and regular graduates in the pilot career field with an odds ratio of 0.471. ABM and CSO regular graduates have similarly low odds of retention in comparison to their DG counterparts with odds ratios of 0.304 and 0.302, respectively.

Finally, Figure 9 shows the odds ratio of retention for officers with dependents. Officers with dependents are set at the baseline value of one. All three career fields show virtually similar odds ratios for retention for officers without dependents. Pilots have the lowest odds with 0.229 while CSO is close at 0.276. ABM has the best odds of retention compared to officers with dependents among the three career fields at 0.355.

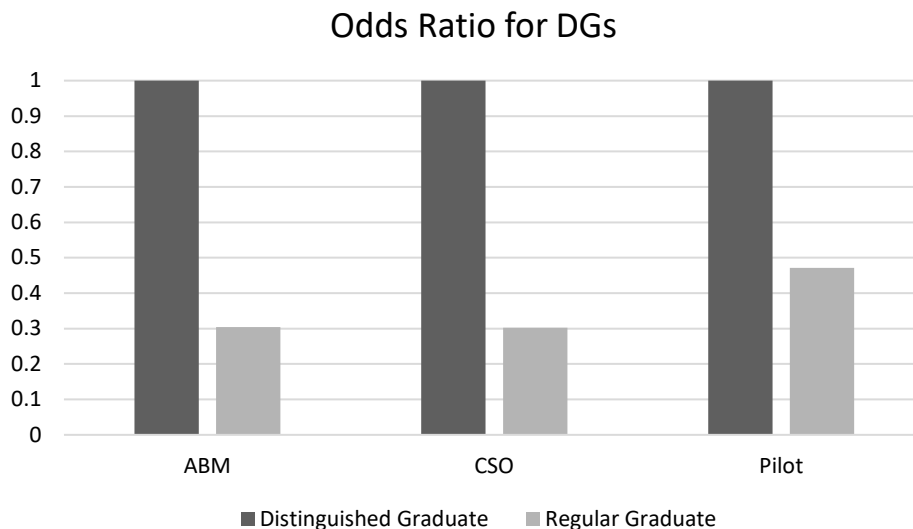


Figure 8. Odds Ratio of Retention for Distinguished Graduates

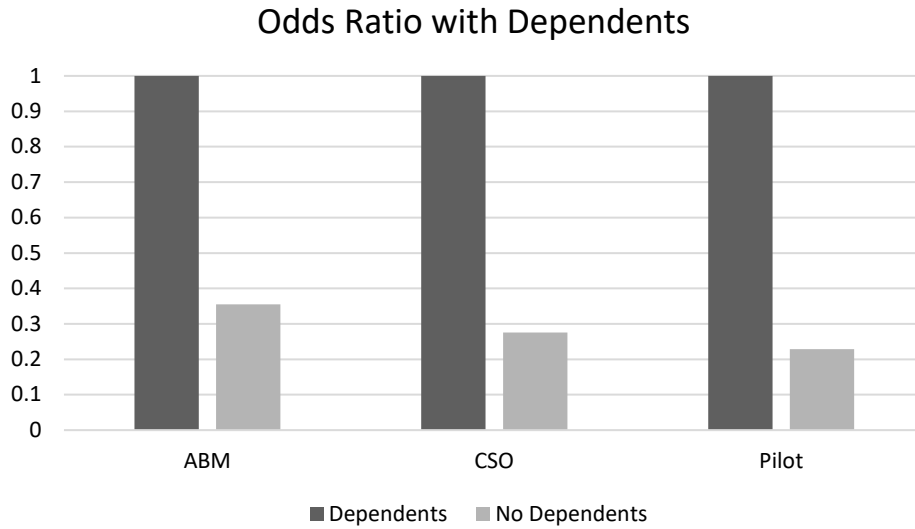


Figure 9. Odds Ratio of Retention for Officers with Dependents

4.2.4. Analysis of ABM CYOS Subset p-Values

For the analysis of the CYOS subsets, significant indicators are discussed in terms of Wald Chi-Square p-values only. Odds ratios are not discussed in the breakdown for two reasons: 1) due to the poor results for the ABM career field, no useful analysis could be conducted within and between career fields and 2) when the data are significant, odds ratio analysis at the CYOS subset level does not produce any additional insight compared to the odds ratio discussion at the career field level. Table 2 summarizes the p-values for each CYOS subset with respect to the model's covariates.

Table 2. p-Values for ABM by CYOS

CYOS	Marital Status	Gender	Commission Source	Prior Service	DG	Dependents	Number of Observations
0 – 6	0.0216	0.1573	0.0431	0.3473	0.1169	0.1545	2,083
4 – 8	0.0047	0.0001	0.0159	0.1214	0.2196	0.0054	1,836
8 – 14	0.0483	0.0132	0.0160	0.0020	0.0913	0.0011	1,161
12 – 19	0.1104	0.8839	0.0029	< 0.0001	0.0109	0.5305	682
20 – 22	0.6539	0.0814	0.9625	0.0807	0.2581	0.9440	499

The first impression from a visual inspection of Table 2 is a general lack of significance. Only the subset, 8-14 CYOS, produced a model that converged according to SAS's convergence criteria. The rest of the CYOS ranges are consumed by a majority of factors that are not significant, which is confirmed by SAS's detection of quasi-complete separation of data points. This lack of significance may be due to the small dataset that characterizes the ABM career field. The initial population was only 2,542 officers for the entire ten-year span and was reduced further by the elimination of truncated observations. Without a rich dataset, correlations between the covariates and the response variable may be difficult to ascertain, and the data that are available may be too noisy. Further study into the ABM career field could be conducted to determine why this career field varies so greatly in its retention factors as compared to the rated cohorts (the pilots and CSOs).

4.2.5. Analysis of CSO CYOS Subset p-Values

Fortunately, a more reliable analysis is obtained when analyzing the CSOs. The Wald Chi-Square p-values for the CSO's CYOS subsets is summarized in Table 3.

Table 3. p-Values for CSO by CYOS

CYOS	Marital Status	Gender	Commission Source	Prior Service	DG	Dependents	Number of Observations
0 – 6	0.0039	0.0019	0.3273	0.0013	0.0245	0.1075	6,446
4 – 8	0.0026	< 0.0001	0.0008	0.9981	0.0001	< 0.0001	5,931
8 – 14	0.0675	< 0.0001	0.0252	0.0065	< 0.0001	< 0.0001	4,278
12 – 19	0.1329	0.0006	0.0002	< 0.0001	< 0.0001	0.1048	3,215
20 – 22	0.6778	0.8449	0.0057	0.0069	< 0.0001	0.4898	2,582

Table 3 results indicate that marital status and dependents are not significant factors within three of the five CYOS subsets (specifically later in a CSO's career). This may be

explained by the timing of career and family life. Marital status is significant at the beginning of the career and only becomes less significant as the officer's career progresses (8 CYOS and beyond). Dependents appear to be significant in the middle of the CSO's career, which coincides with the traditional timing of family formation. As the career progresses, the presence of dependents ceases to be significant (12 CYOS and beyond). Gender stops being a significant factor at the end of the career (20-22 CYOS) while commissioning source and prior service appear to have no significance in the initial portions of an officer's career. This could be due to ADSC commitments for the commissioning source and total time in service for the prior service members.

4.2.6. Analysis of Pilot CYOS Subset p-Values

Logistic regression at the CYOS subset level also provides greater retention model fidelity with respect to the pilots. Table 4 summarizes the p-values for the pilot CYOS breakdown.

Table 4. p-Values for Pilot by CYOS

CYOS	Marital Status	Gender	Commission Source	Prior Service	DG	Dependents	Number of Observations
0 – 6	0.0006	0.0003	0.0502	< 0.0001	0.0779	0.3150	21,531
4 – 8	< 0.0001	< 0.0001	0.1590	0.1070	< 0.0001	< 0.0001	19,416
8 – 14	< 0.0001	< 0.0001	0.2370	0.0457	< 0.0001	< 0.0001	14,046
12 – 19	0.0024	0.2226	0.0026	< 0.0001	< 0.0001	< 0.0001	8,562
20 – 22	0.0466	0.9972	< 0.0001	0.0905	< 0.0001	0.9373	6,227

The two most prominent factors in Table 4 are gender and commissioning source. Gender becomes an insignificant factor in the later part of a pilot's career (12-19 and 20-22 CYOS) while commissioning source is not significant in the initial phase of their career (prior to 14 CYOS). Dependents are not a significant factor for retention at the very

beginning or end of a pilot's career. In the beginning, pilots are in training and not deploying. Their work-life ratio is likely not impacted by their military service. As they progress in their career, the demands of the Air Force increase simultaneous with the needs of the pilot's family. They cannot focus on their career without sacrificing some of their connection to their family. At the end of their career, the children are older and/or the pilots that remain to this point are the ones with families that have acclimated to the poor work-life ratio. Additionally, those with prior service have a higher total active federal military service (TAFMS) than their CYOS indicates. This may mean that a majority of those with prior service have already separated by the 20-22 CYOS and are now in the minority and therefore, less significant to retention patterns.

4.3 Survival Analysis

4.3.1. Introduction

Personnel Analysts use sustainment curves as part of their analysis efforts. These curves portend to provide retention probabilities. Survival function curves are examined as an alternative to the current sustainment lines.

Survival analysis is utilized to estimate survival function curves and parameter estimates for significant factors for officer retention. Nonparametric methods such as Cox proportional hazards regression are powerful tools because the techniques do not require the data to be normally distributed and can compensate for issues with censoring (observation period does not include all time events for some observations in the dataset) [51, 35]. SAS code for survival analysis is available in Appendix B.

4.3.2. Data

During the data refinement portion of the analysis, probability and cumulative distributions functions, frequency tables, and simple statistical calculations (mean, variance, standard deviation, etc.) are assessed to gain an impression of the structure and patterns of the data. Empirical inspection shows that the data do not follow a normal distribution and the predominant population characteristics for all three career fields are married, male, with dependents, no DG status, no prior service, and commissioned through ROTC or USAFA.

Assumptions made for this analysis include compensating for potential left-censoring of the data. Left-censoring occurs when data exists but is not observed prior to the initiation of the observation period [51]. The short timeframe for the data extracts ensures that some observations will not encompass the start of an officer's career. Therefore, this analysis assumes that if an officer is a commissioned officer at the time that they are observed (2006-2015), then that officer was an officer starting at time zero of their career. For example, if a CSO is an officer with 4 CYOS in 2006, then it can be assumed that the CSO was commissioned as an officer in 2002 and their career started at 0 CYOS. This portion of the officer's career is outside of the observation period, but because of this assumption, the officer's data may remain within the analysis and is considered left-censored.

Right-censored data is excluded from the analysis (producing smaller dataset populations than for logistic regression). An example of right-censoring is an officer who begins their career in 2010. The officer will only have 5 CYOS by the end of the observation period (2015) and may not have separated from military service by this point.

It can be assumed that the officer's career will end, but the time of that terminal event has either occurred and is outside of the observation period or has not occurred yet and is currently unknown. While survival analysis techniques can accommodate for right-censoring using estimation methods to predict terminal events for unobserved data, this exclusion of right-censored data will ensure that there is no skewing of the results due to the limitations in accuracy of these estimation methods [51].

For logistic regression, the analysis was broken down by CYOS subsets; however, for survival analysis, the breakdown is at the career field level. Because the survival function curves are a graphical depiction of instantaneous attrition rates at given times in the span of a 20-year career, examining the CYOS subsets would not provide additional information. It would be equivalent to zooming in on a portion of the survival function curves created for the overall population, but that data is already available at the career field level. Therefore, the analysis remains at the career field level and is analyzed with respect to the significant factors for retention from the logistic regression and Cox proportional hazard regression models.

4.3.3. Kaplan-Meier Survival Functions

Kaplan-Meier curves are produced with SAS's *proc lifetest* command. This technique produces "survival function estimates to assess proportional hazards for categorical covariates" [39]. A depiction of the difference in survival curves for pilots with respect to marital status can be seen in Figure 10 (shaded regions around the curves represent the 95% confidence band). Marital status 0 is single, 1 is married, and 2 is previously married (divorced, widowed, annulled, etc.). Within all three career fields, the retention curves for married and previously married officers crossed at various points in

the initial phases of the officers' careers (Kaplan-Meier curves discussed in this section but not shown can be referenced in Appendix C). Married officers became the dominant retention curve around 12-15 years for all rated officers. Single officers had consistently low retention curves with the largest slopes seen at the end of their respective initial training ADSCs. After the initial loss, all career fields saw a relative plateau in retention. For example, in Figure 10, the slope steepens at approximately 8 CYOS (going from ~80% retained to ~20% retained) and continues till 12 CYOS, where it plateaus till 20 CYOS.

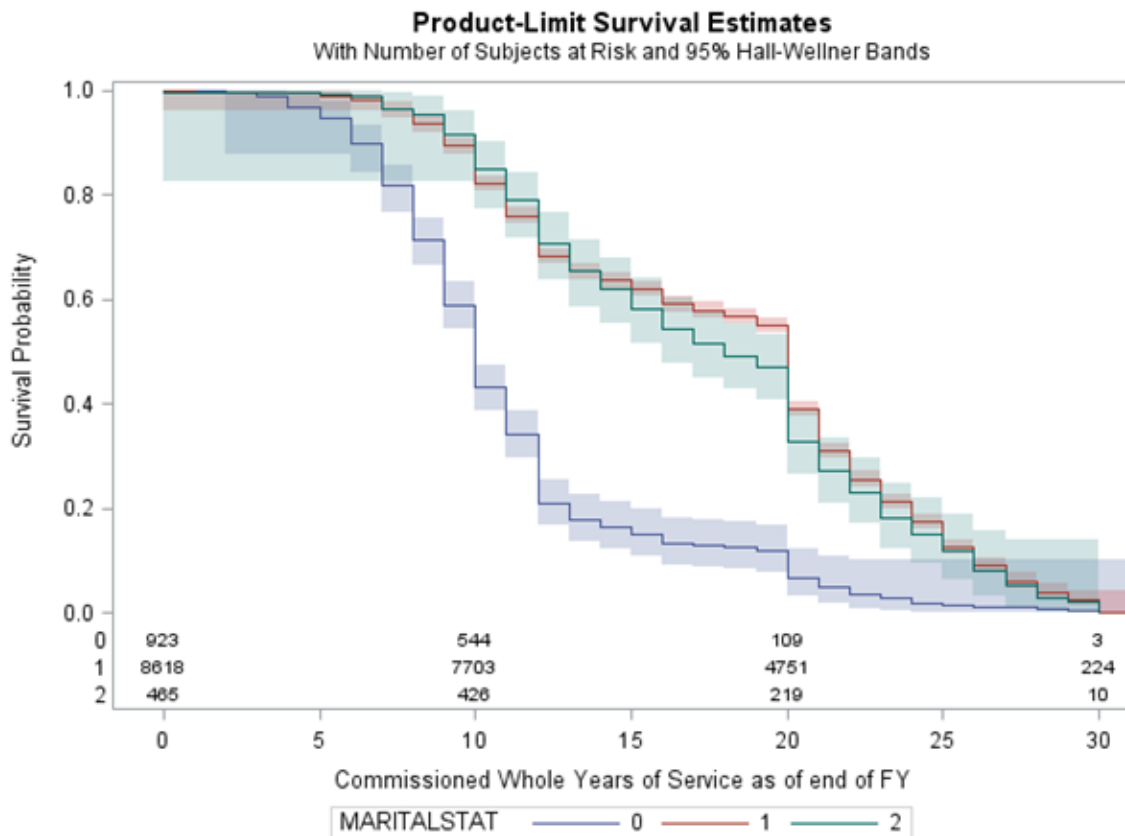


Figure 10. Kaplan Meier Curve for Marital Status (Pilot)

In the odds ratios results from logistic regression, the odds of retention for females within the CSO career field were at the lowest rate compared to ABMs and pilots.

Figure 11 shows the pronounced difference in retention rates between the genders with a majority of the attrition occurring at the end of the initial training ADSC. ABM and pilot curves are provided beneath the CSO curve for comparison. Additionally, ABM odds ratio of retention were the closest between males and females, and this is reflected in the KM curve provided.

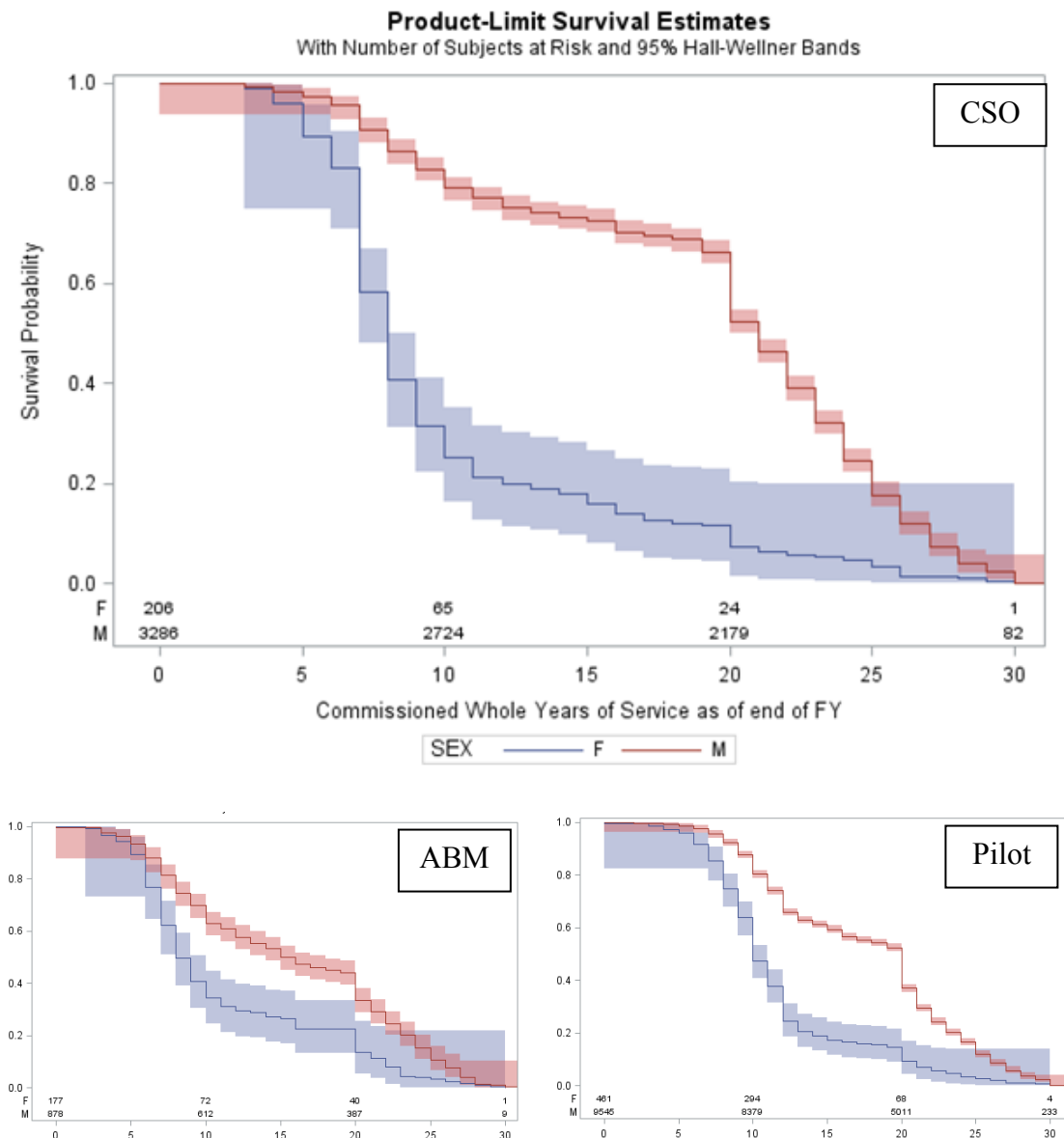


Figure 11. Kaplan-Meier Curve for Gender

Figure 12 illustrates the survival curve estimates for ABMs with respect to source of commissioning (SOC). SOC 1 represents Other commissioning sources (e.g., Army OCS), 2 is USAFA, 3 is ROTC, and 4 is OTS. The KM estimates for ABMs were worth noting with respect to this variable for two reasons. First, the population for the other commissioning source was so small that the 95% confidence bands for this result extends almost the entire vertical field, meaning that there is virtually no accuracy to that element of the data. Second, the odds ratios for retention with respect to commissioning source was largest with the USAFA ABMs, and Figure 12 corroborates that result. While the KM curves for pilots are tightly banded and cross each other at various points with no dominant result, the CSO curves are not as distinct from one another, and at 10 CYOS, USAFA graduates become the dominant retention curve, mirroring the odds ratios results.

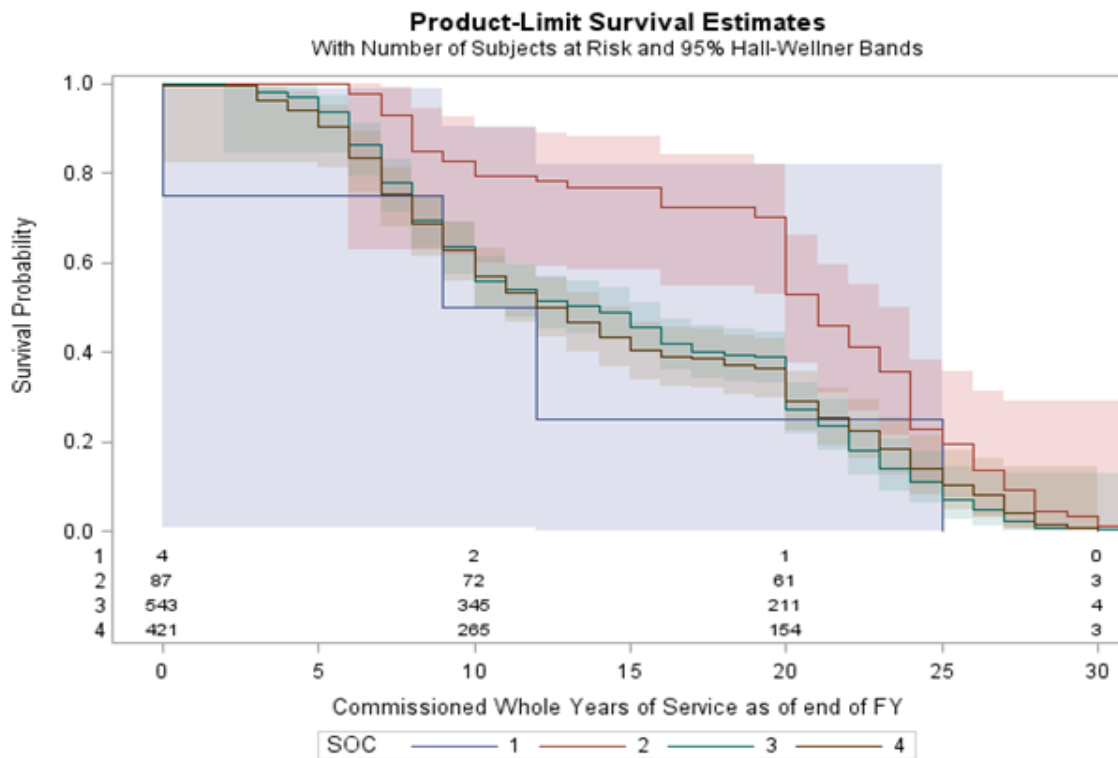


Figure 12. Kaplan-Meier Curve for Commissioning Source (ABM)

The effect that prior service has on retention is captured in Figure 13. The curve for ABM officers is arbitrarily selected to illustrate this variable because all three career field's KM curves look virtually identical with the crossover pattern in the initial years of the career. This exchange of dominance ends at year 12-15, and the retention rate of officers with no prior service becomes dominant.

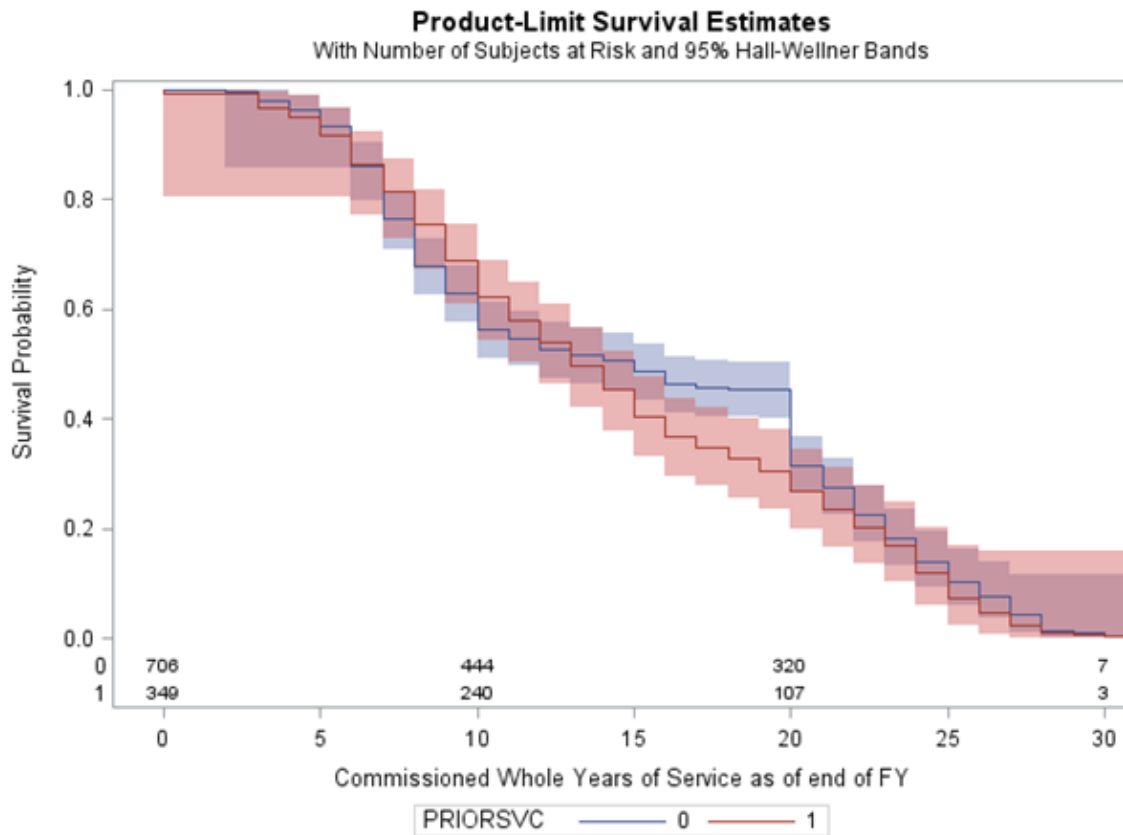


Figure 13. Kaplan-Meier Curve for Prior Service (ABM)

Figure 14 illustrates the KM curve with respect to DG status for pilots. The odds ratio results showed that DG status was the least significant in retention for pilots. This result is confirmed in Figure 14. CSO and ABM curves are provided for comparison. Retention for regular graduates is relatively close to DG for pilot and CSO; however, with

ABM officers, retention for regular graduates sees a sharper decline in years 5-10. Rates plateau at ~40% for regular graduates compared to the ~65% retention rate of DGs. DG CSOs maintain a strong retention rate throughout a 20-year career with a rate above ~80%. Conversely, retention rates for DG pilots decrease from ~95% to ~75% starting at year 8 (end of ADSC).

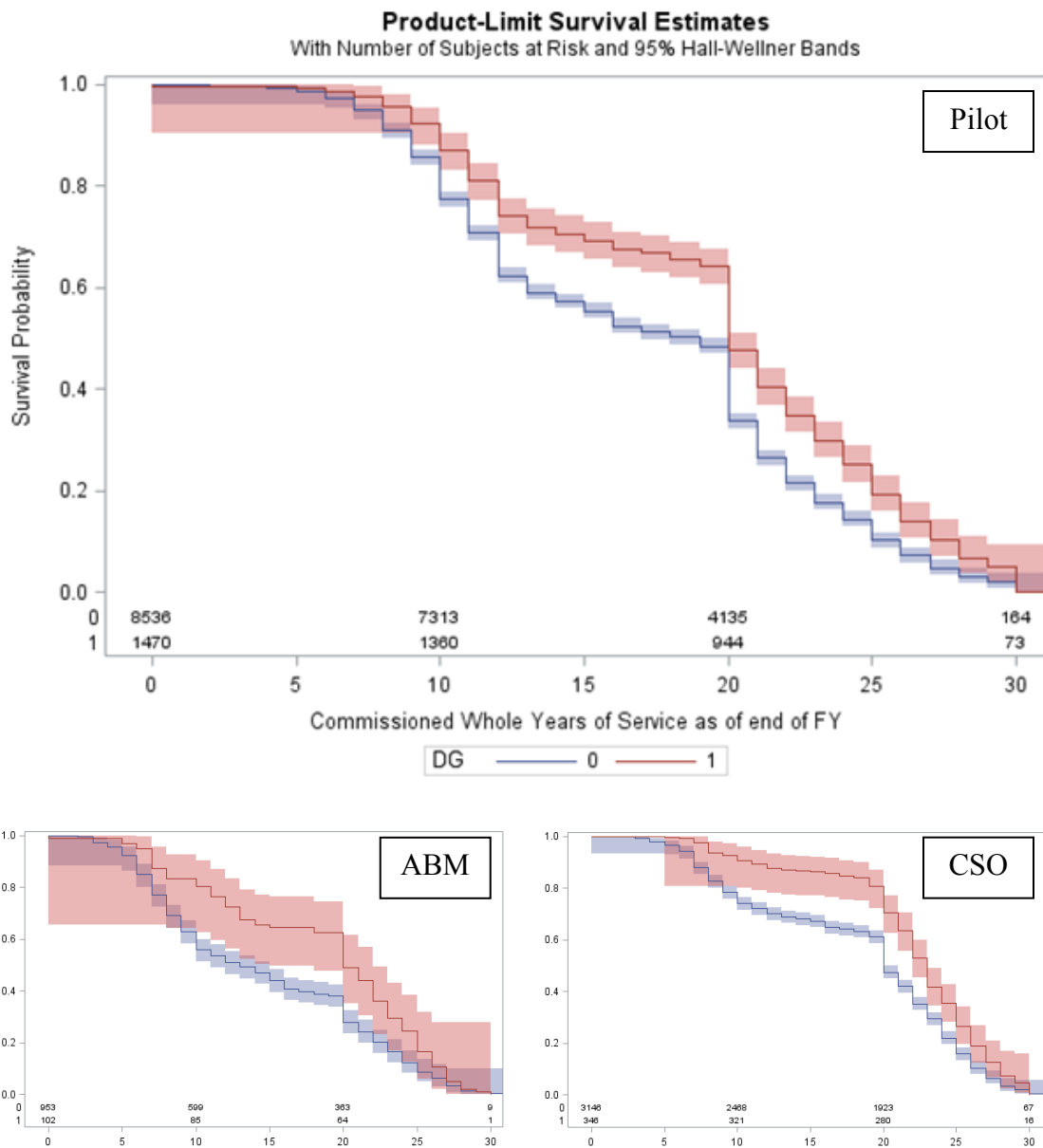


Figure 14. Kaplan-Meier Curve for DG

Finally, the effect on retention rates with respect to whether an officer has dependents is examined in Figure 15. All three KM curves have similar structure and characteristics. CSOs are shown in this figure because they exhibit the highest retention rates for officers with dependents. The probability for survival stays above ~75-80% until the 20-year point. Pilots plateau at ~60-65% between 10-20 years. ABMs enter into the 20-year mark at the lowest retention rate for officers with dependents at ~50% retained.

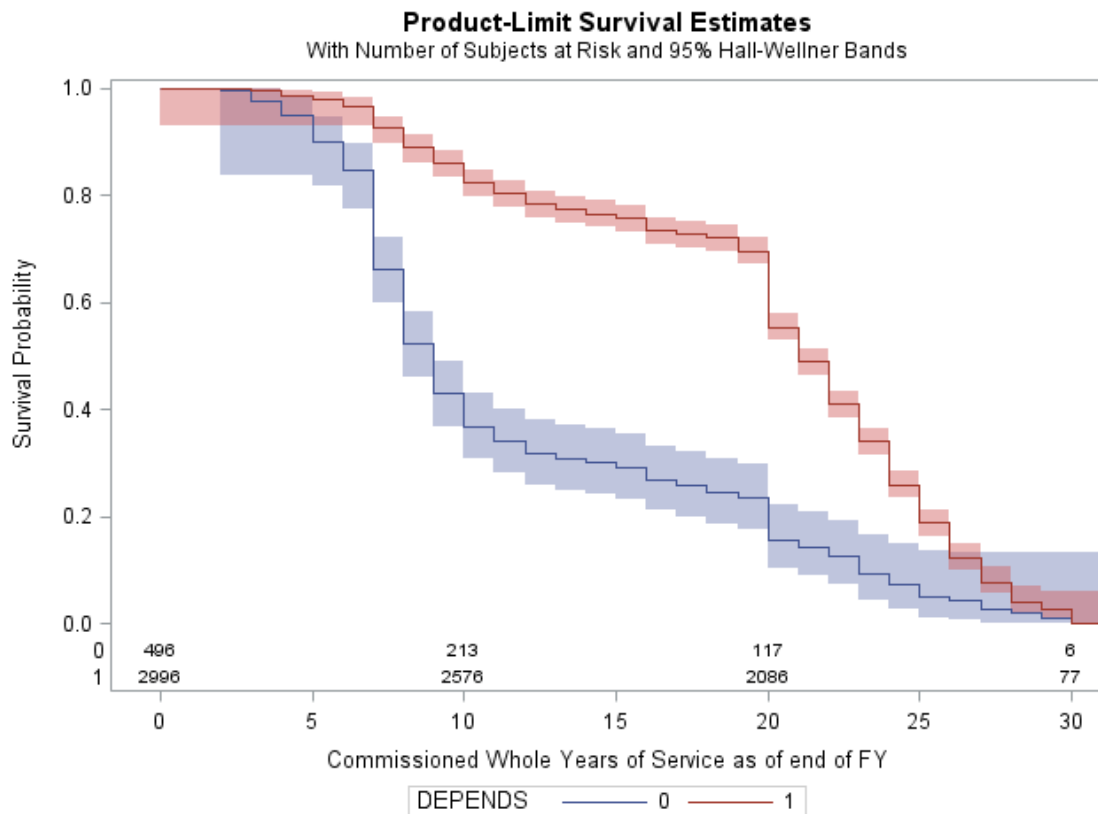


Figure 15. Kaplan-Meier Curve for Dependents (CSO)

4.3.4. Cox Proportional Hazards Regression

Survival analysis is performed in SAS using the *proc phreg* command, which produces the Cox proportional hazards regression model. This regression technique is

useful because “estimating predictor effects does not depend on making assumptions about the form of the baseline hazard function. . . . Instead, we need only assume that whatever the baseline hazard function is, covariate effects multiplicatively shift the hazard function and these multiplicative shifts are constant over time” [39].

Cox PHREG results are consistent with the results obtained with logistic regression. All three of the career fields produce models that converged under SAS’s relative gradient convergence criterion. Additionally, all three models produced stronger model fitness scores with the covariates included in the model as opposed to the model with only the intercept. Table 5 summarizes the Wald Chi-Square p-values for the covariates for each of the career fields (significance level of 0.05). All variables are significant factors for officer retention within all three career fields. Table 6 shows the maximum likelihood estimates generated by the Cox PHREG model.

Table 5. Cox PHREG Covariate p-Values for Pilot, CSO, and ABM

Covariates	Pilot	CSO	ABM
Gender	< 0.0001	< 0.0001	0.0025
Prior Service	< 0.0001	< 0.0001	0.0035
DG	< 0.0001	< 0.0001	0.0009
Dependents	< 0.0001	< 0.0001	0.0002
Marital Status	< 0.0001	< 0.0001	< 0.0001
Commissioning Source	< 0.0001	< 0.0001	< 0.0001

Table 6. Cox PHREG Model Parameter Estimates

	ABM	CSO	Pilot
Gender (female)	0.27489	0.93982	0.48336
No Prior Service	– 0.22482	– 0.22317	– 0.21002
Regular Grad	0.35134	0.36651	0.34538
No Dependents	0.46810	0.50125	0.59054
Single	0.63586	0.49493	0.58627
Married	0.11861	0.09041	0.11539
SOC (Other)	0.25872	0.34459	0.33668
SOC (USAFA)	– 0.45282	– 0.22091	0.06184
SOC (ROTC)	0.12069	0.14813	0.18874
Number of Observations	1,055	3,492	10,006

Finally, Figure 16 illustrates the range of the survival function curves for all possible combinations of covariate settings for the officer populations (e.g., one combination of covariate settings is female, single, no dependents, DG, no prior service, USAFA). The most interesting result is the tighter banding of the pilot population in the initial and final phase of their career compared to ABMs and CSOs. This is likely due to the longer ADSC that restricts their opportunities for separation till after the 10-year mark, resulting in a smaller timeframe for attrition compared to ABMs and CSOs.

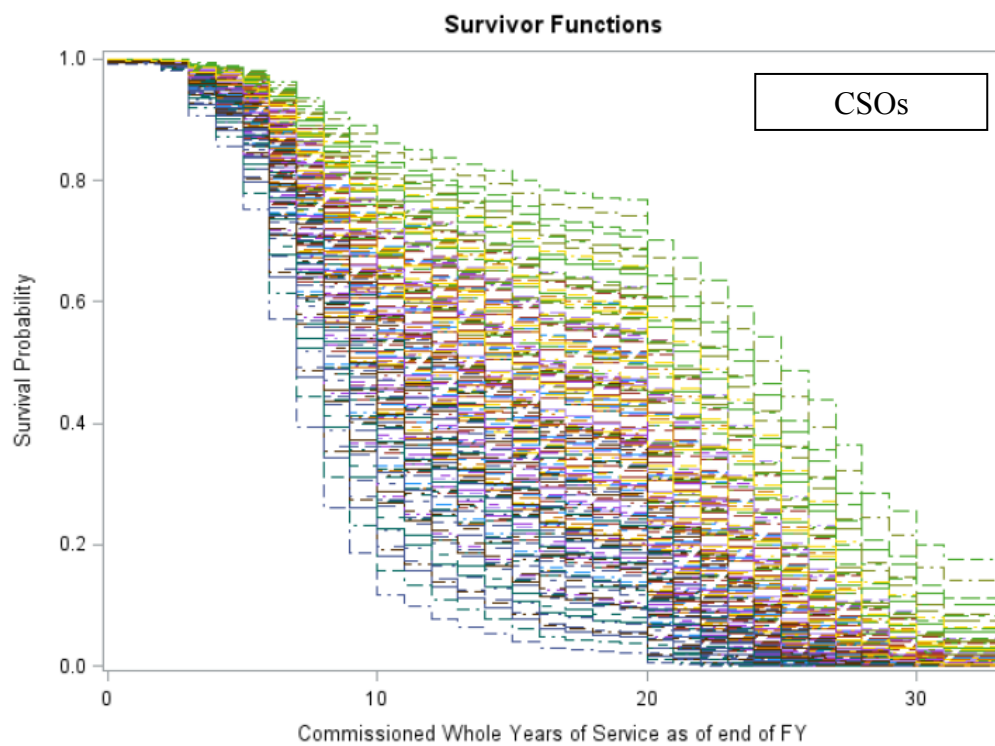
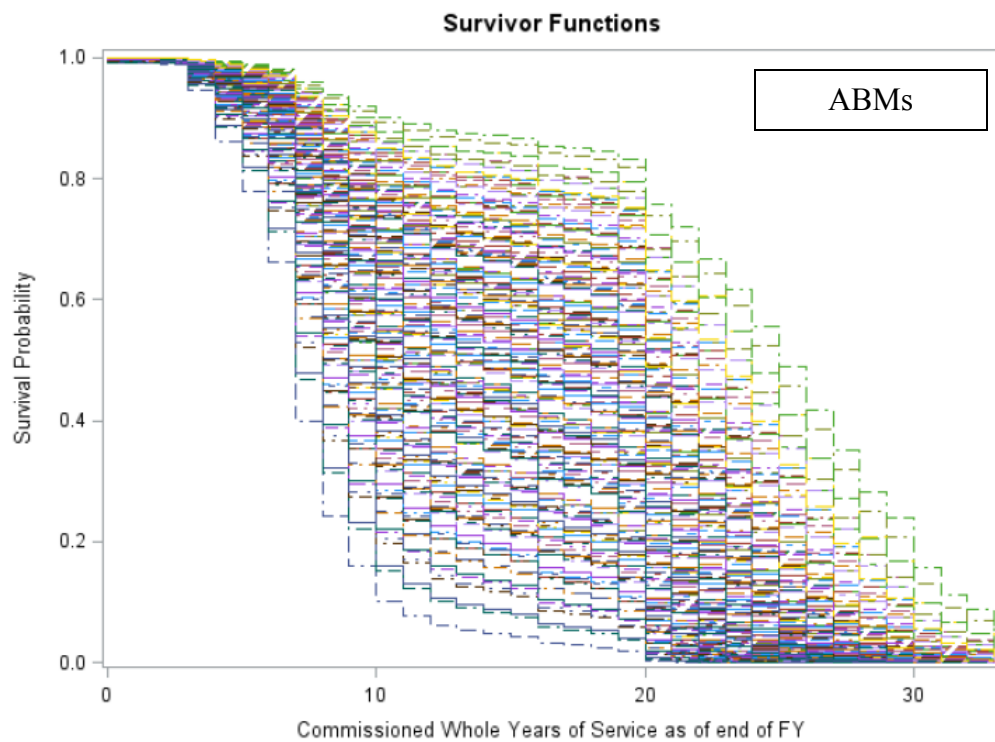


Figure 16. Range of Survival Functions

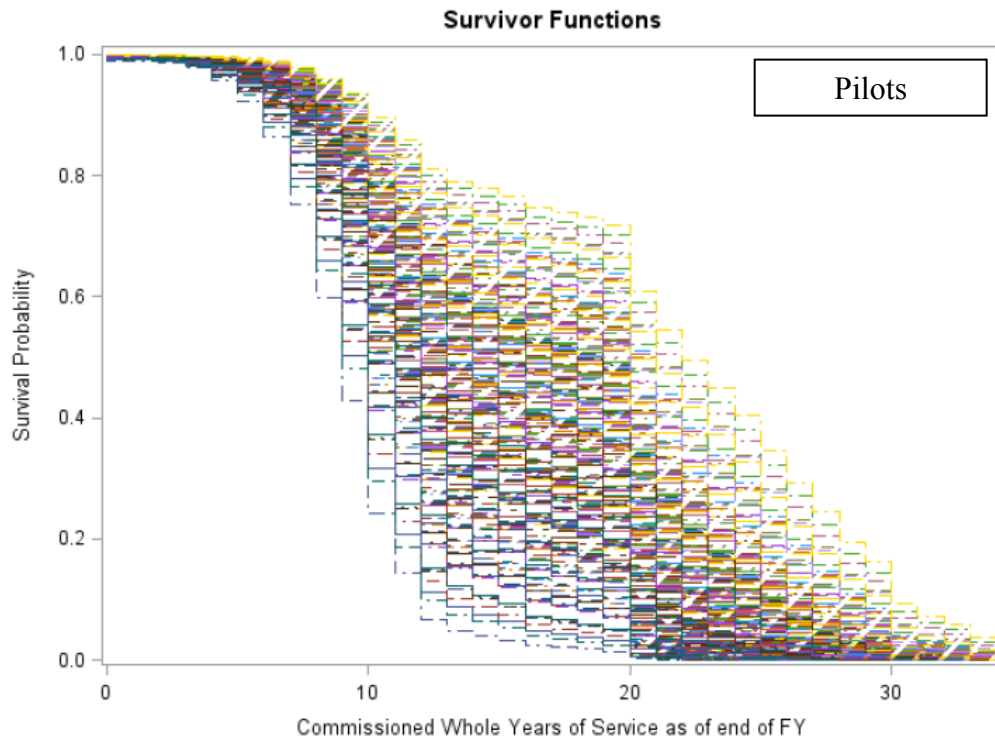


Figure 16 (Continued). Range of Survival Functions

The Cox PHREG model determined that all variables are significant, which equates to 196 possible survival functions for each career field. Not all 196 combinations may actually exist within these populations, but even with a reduced number of possible combinations, there are still too many to discuss each in detail. Instead, to illustrate retention trends, the overall population curves captured by the KM curves were discussed in Section 4.3.3. The key value in the Cox PHREG models is the ability to pinpoint subpopulation retention patterns and focus on factors that may be significant to retention in these subpopulations, which is not possible in current HAF/A1 modeling processes.

V. Results and Analysis

5.1 Introduction

Logistic Regression and Survival Analysis have generated comparable products to the sustainment lines created by HAF/A1. This section compares survival curves to the sustainment lines for each career field and then explores the potential influence of economic and political factors.

5.2 Results

HAF/A1 creates monthly products that summarize career field health (CFH) for all Airmen with respect to total Air Force and broken down by AFSC. Figure 17 shows the CFH chart for Pilots (11X) current as of June 2016 [52]. These data are created using historical data from the past five years. The red line is the sustainment line generated by HAF/A1 under their current calculation process with an additional “smoothing” process. This line is generated once a year and is reused in each monthly CFH chart release. For comparison, this chart is overlaid with the KM curves for the Pilot career field created by the survival analysis (blue staggered line).

With the exception of the initial years in Figure 17 (initial years represent attrition during training which was excluded from this research but could easily be incorporated in the survival analysis) and the smoothing done by HAF/A1, the lines created by survival analysis and the sustainment line generated by HAF/A1 have similar proportions and shapes. Retention rates decrease prior to the ten-year milestone in both curves (both showing ~80% retention at ten years). The same knee at the 20-year mark is mirrored in

both with approximately 50% retention at this point and a steep decline in retention as many pilots choose to retire.

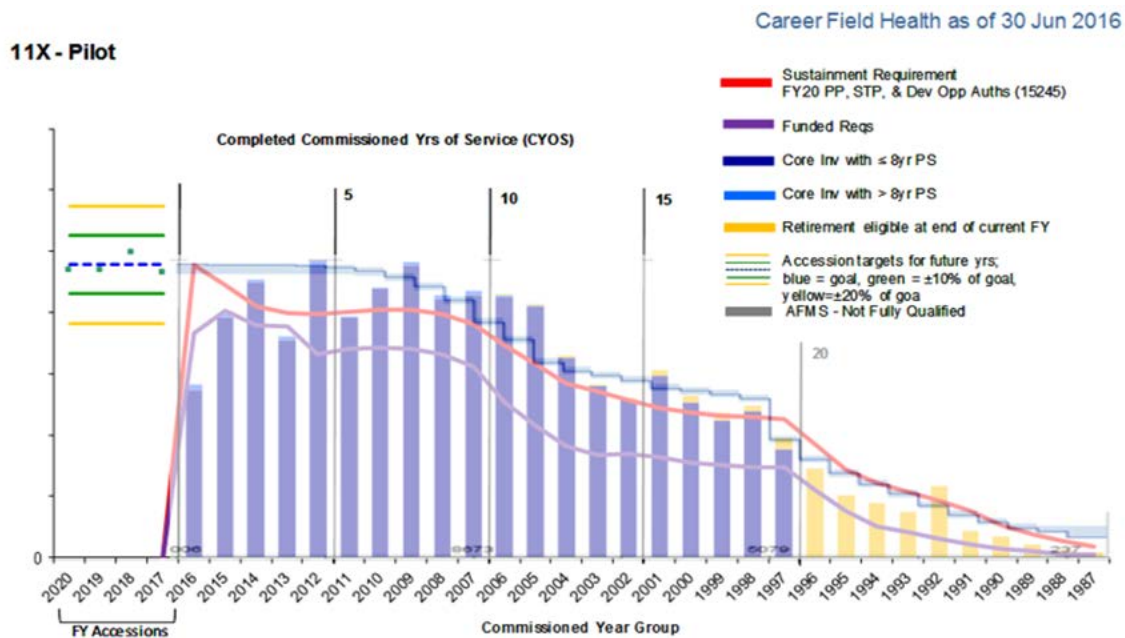


Figure 17. HAF/A1 Sustainment Line for Pilot with KM Curve Overlay

Figure 18 shows HAF/A1's sustainment line and the survival function curve for CSOs. Just as with pilots, there are similar shapes but proportions are slightly higher in the plateau for the CSO KM curve compared to the sustainment line. The rate of attrition begins to increase at around the five-year mark. The increased attrition rate is higher in the sustainment line. This difference could be due to both the added retention data (ten years versus five years) and the smoothing done by HAF/A1. Pilots plateau at about the 55% range for the ten-year period before 20 years; however, you can see the higher retention rate at the plateau point for CSOs in both curves (~60-65%). Finally, CSO retention lines have a sharp cliff at the 20-year milestone. An approximately 60% retention drops to ~20% within five years after CSOs reach retirement.

12X - Combat Systems Officer

Career Field Health as of 30 Jun 2016

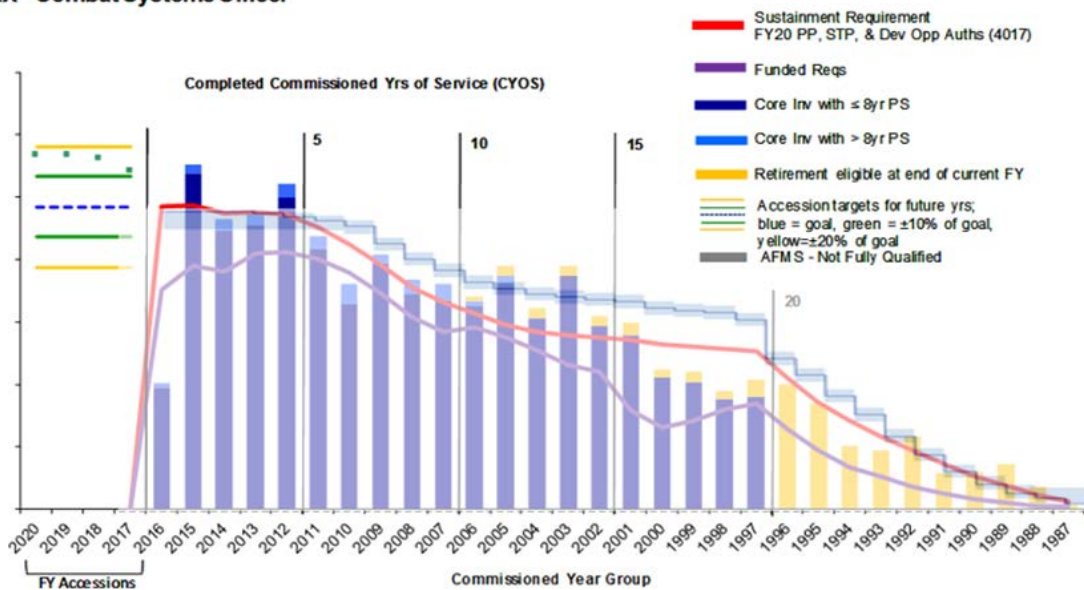


Figure 18. HAF/A1 Sustainment Line for CSO with KM Curve Overlay

Finally, Figure 19 shows the sustainment line with survival curve overlay for ABMs. Again, a decline in retention rates starts just prior to five years in both graphs (retention is approximately 95% by five years). In the survival curve, retention rates decrease from five to ten years and then plateau at ten with retention rates flat-lining at approximately 40-50%; however, the plateau in the HAF/A1 line is approximately 50-60%. The 20-year cliff can be seen, and retention from 40% to 10% by year 25 is reflected in both products. Of the three career fields, ABM's fit to the HAF/A1 survival curves is not as strong as the other two career fields. This may be due to information gleaned from the extension of the historical data (ten years of extracts versus five years of historical data). Overall, survival analysis appears to give comparable products to HAF/A1's current sustainment line calculations.

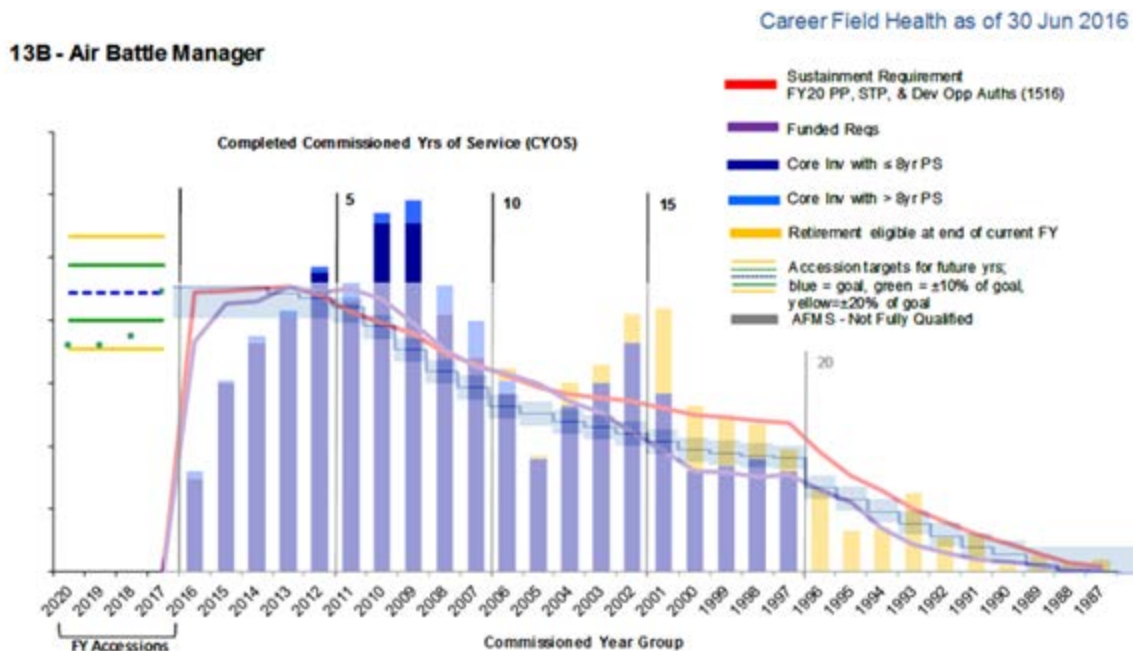


Figure 19. HAF/A1 Sustainment Line for ABM with KM Curve Overlay

5.3 Analysis of Model Adequacy

To examine lack of fit of the model created by the Cox Proportional Hazards regression analysis, the Martingale Residual Plot and Deviance Residual Plot are examined for all three career field models. Figure 20 and Figure 21 show the Martingale and Deviance residuals, respectively, for the pilot model. Data is skewed to the top because of the single event setting of Cox model. Patterns in the Martingale residuals suggest that continuous variables are not properly fit [53]; however, all variables in this model are categorical, so this is an expected result. There are one or two large values in the -15 to -20 range of the Martingale axis that suggest potential issues.

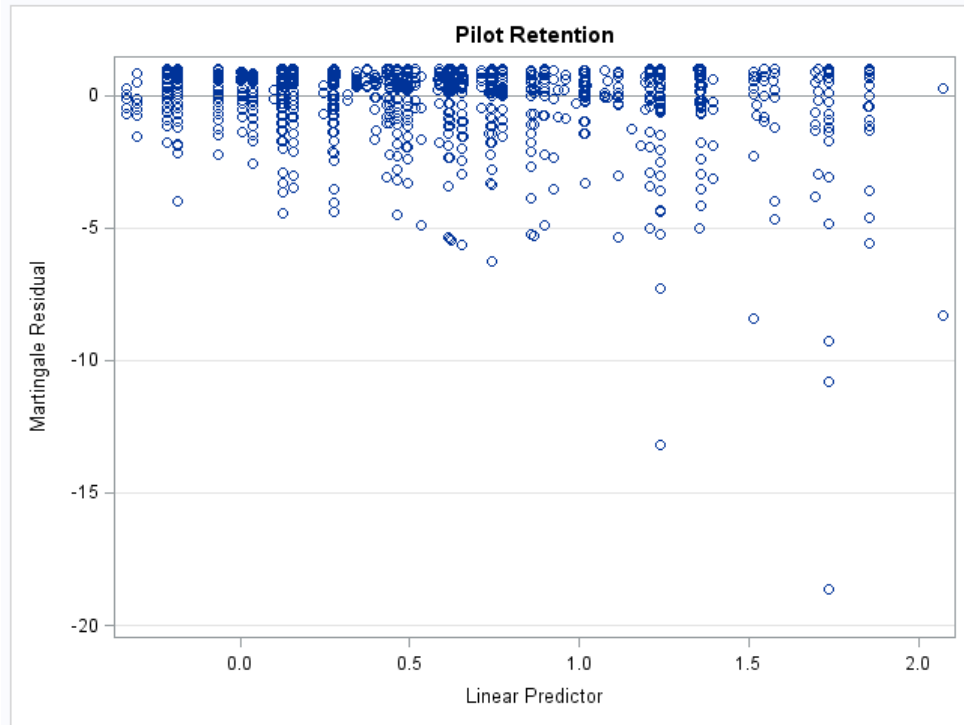


Figure 20. Martingale Residual Plot for Cox Proportional Hazards Model (Pilot)

Examination of the Deviance residuals in Figure 21 suggests that perhaps a few of the data points are, in fact, outliers. Further investigation is required into these data points to confirm if they should be dismissed from the dataset. Currently, there appears to be no issues with lack of fit in the model with individual observations.

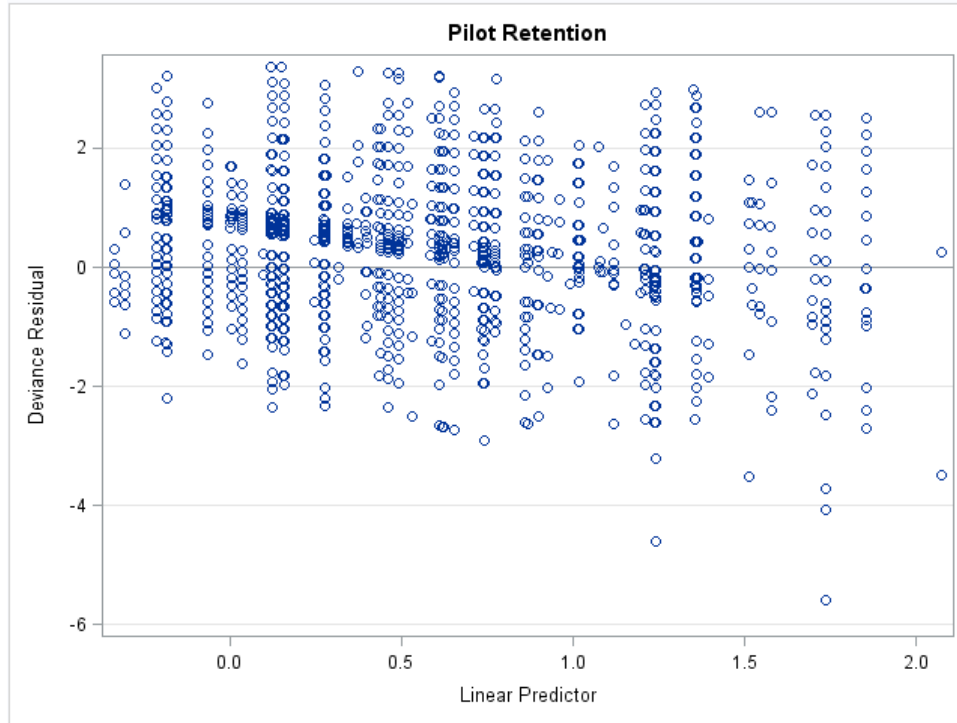


Figure 21. Deviance Residuals Plot for Cox Proportional Hazards Model (Pilot)

Figure 22 and Figure 23 show the residual plots for the CSO survival model. Again, potential indications of large values can be seen in the Martingale residuals, but inspection of the Deviance residuals dismisses the potential for outliers in this model. The Deviance residuals in Figure 23 show a relatively even distribution of the observations on either side of the zero line.

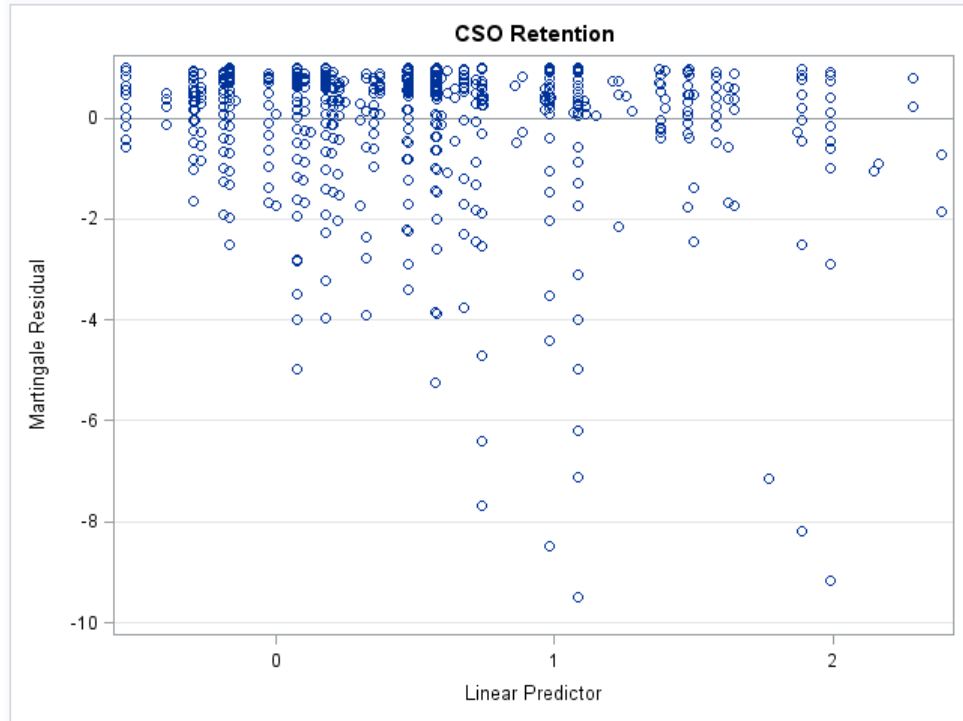


Figure 22. Martingale Residual Plot for Cox Proportional Hazards Model (CSO)

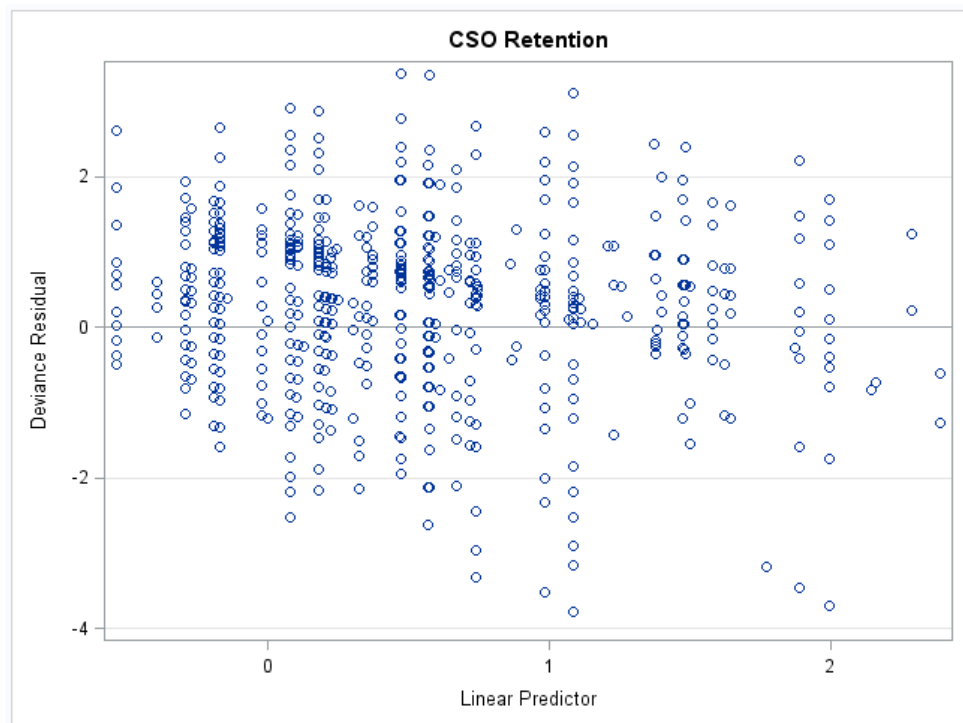


Figure 23. Deviance Residual Plot for Cox Proportional Hazards Model (CSO)

Finally, the Martingale and Deviance residual plots for the ABM survival model is shown in Figure 24 and Figure 25. The Martingale residuals do not show any excessively large numbers and the even distribution of the Deviance residuals suggest no concern with outliers. Overall, no issues with model adequacy exist for any of the three models.

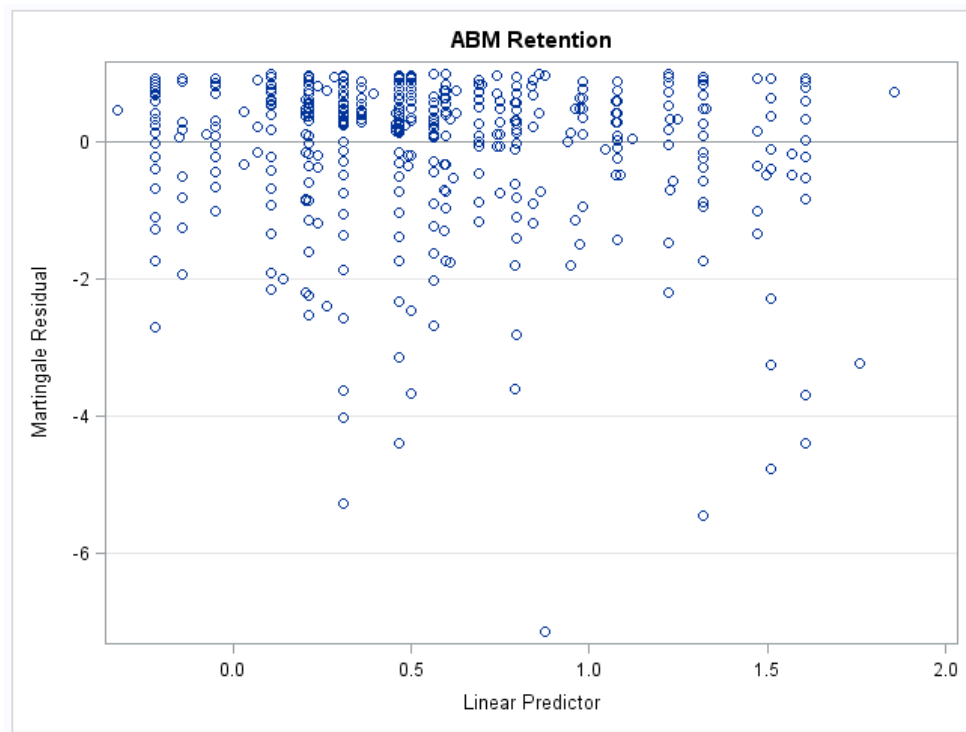


Figure 24. Martingale Residual Plot for Cox Proportional Hazards Model (ABM)



Figure 25. Deviance Residual Plot for Cox Proportional Hazards Model (ABM)

5.4 Analysis of Economic and Political Factors

An extension to the survival analysis conducted in Chapter 4 incorporates economic and political factors. Just as with the demographic data, a stepwise survival analysis using SAS's *phreg* command is implemented to determine which economic or political covariates should be left in the survival model to best explain retention rates for rated officers.

The stepwise regression examines six demographic variables (marital status, gender, commission source, prior service, distinguished graduate, and dependents), two political variables (male-to-female ratio and deployment rate), and four economic variables (unemployment rate, PMI, CCI, and durable goods new orders report number). Of these

variables, the six demographic variables are reconfirmed as significant to rated officer retention among all three rated career fields; however, only one economic or political factor may also be significant to retention among all three career fields. Model adequacy was confirmed using residuals plots. Table 7 shows the p-values for each career field with respect to the variables that were incorporated into the model with the stepwise method. The CCI variable is not listed in the table because it was not statistically significant to retention by any of the career fields. In addition to the six demographic factors, three economic/political factors may be significant to pilot retention. These factors are male-to-female ratio, unemployment rate, and the PMI value. For CSO retention, male-to-female ratio and PMI may also be statistically significant. ABM retention appears to be the only career field that may be impacted by deployment rates. Among all three career fields, the values associated with new orders for all manufacturers of durable goods was found to be significant to the retention model.

Table 7. Stepwise PHREG Model p-Values

Covariates	Pilot	CSO	ABM
Marital Status	< 0.0001	< 0.0001	< 0.0001
Gender	< 0.0001	< 0.0001	0.0026
Commissioning Source	< 0.0001	< 0.0001	< 0.0001
Prior Service	< 0.0001	< 0.0001	0.0024
DG	< 0.0001	< 0.0001	0.0026
Dependents	< 0.0001	< 0.0001	< 0.0001
Male-to-Female Ratio	< 0.0001	< 0.0001	--
Unemployment Rate	< 0.0001	--	--
DGR (Total Manufacturing New Orders)	0.0026	0.0168	0.0033
Purchasing Managers' Index	0.0073	< 0.0001	--
Deployment Rate (Career Field Specific)	--	--	0.0002
<i>Note: '--' indicates that the variable was not significant for that career field.</i>			

While only one factor is significant for all three career fields, analysts are frequently more concerned with the retention rates of specific career fields. Economic and political factors that may be significant in retention of specific career fields could also be used to tailor accession or retention programs. The survival model created in this analysis suggests that the male-to-female ratio may be significant for retention rates with respect to pilots and CSOs; however, this is likely not a useful metric as it is neither forward-leaning nor actionable. Examining the effect of a higher female population to retention rates shows that an increase in the proportion of females to males correlates to a decrease in retention rates. Figure 26 shows this relationship.

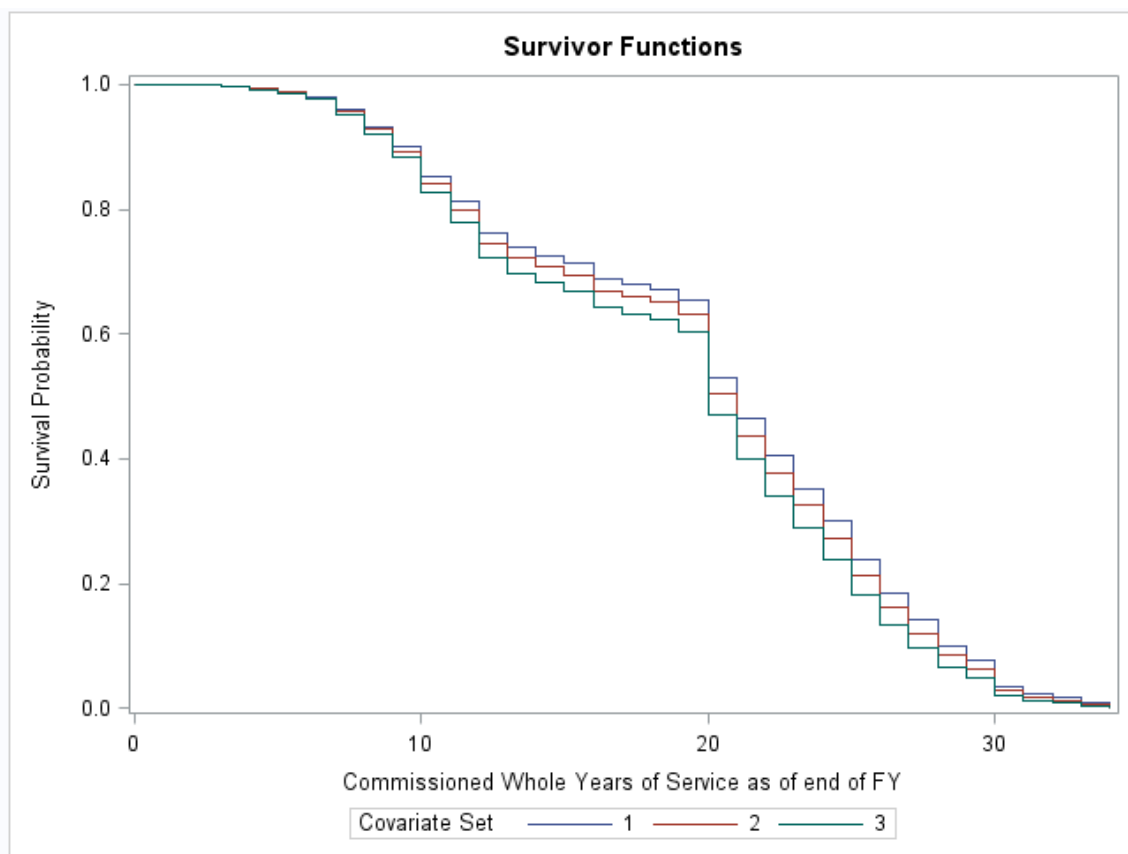


Figure 26. Survival Curve for Pilot and CSOs with Low, Average, and High Female Percentage

Covariate Set 1 in Figure 26 represents a low setting of 18.03% for female-to-male ratio while all other covariates are held constant. Covariate Set 2 is the average (18.93%) and Covariate Set 3 is a high female proportion of 20.36%. The relationship here is direct. Increasing the number of females in the population directly correlates to increasing the number of attritions. Therefore, while the male-to-female ratio may be significant to retention of pilots and CSOs, this result merely reaffirms previous statements about the underlying issues of female retention and the need for better retention policies for female members.

Three other covariates that may be significant to a specific career field's retention rate are the unemployment rate, the PMI, and ABM deployment rate. According to exit surveys conducted by the Air Force, one of the top reasons for attrition of pilots is a high perstempo, i.e., high deployment rate [2]; however, ABM is the only career field where the deployment rate appeared to be significant with respect to retention. Furthermore, the PMI, which is an economic index created in the hopes of acting as an indicator of economic health, is only significant for Pilots and CSOs. The third covariate is unemployment rate. As previously discussed, this indicator is complex in nature. Some economic analysts view it as a lagging indicator because job loss is usually reactive to economic downturns [43]; however, the link between pilots and unemployment rate may likely be tied to airline hires. As there are no direct civilian equivalents to the jobs performed by CSOs and ABMs, they do not see the same relationship between the unemployment rate and retention.

Finally, the relationship between the DGR's New Orders values and retention rates among all rated officers is shown in Figure 27. The New Orders value is an objective measurement based entirely on the domestic manufacturing industry's new orders of

durable goods for the next month. When this value is high, the economy is generally doing well. When this value is low, the economy is not as strong. With all covariates in the survival function held constant at their average value, the New Orders covariate was varied from low, middle, and high values to see the effect on retention.

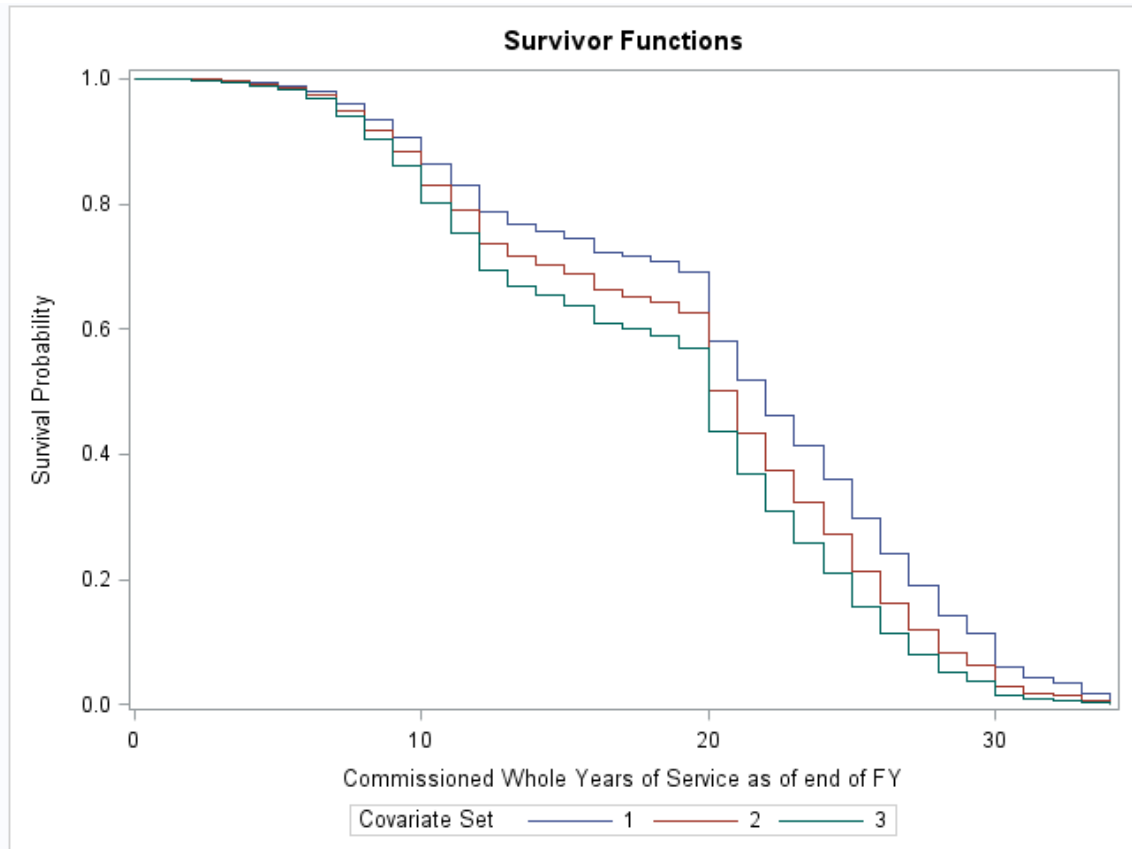


Figure 27. Survival Curve for Rated Officers with Low, Average, and High New Orders Values

As expected, Figure 27 shows that when New Orders are low at \$331,256 million (Covariate Set 1), retention of rated officers is higher with the survival probability increasing by ~5-7% from the average. Average value is represented by Covariate Set 2

(\$446,269 million). When New Orders are high at \$548,071 million (Covariate Set 3), the retention rate for rated officers increases by ~7-10% from the average.

HAF/A1 continually analyzes and adjusts their annual accession and retention goals by fine-tuning their models on a monthly basis. As the Advance Durable Goods Report is a monthly product that reflects future economic health, it shows potential as a useful tool for personnel analysts. When the analysts are fine-tuning their accession and retention programs, they could consult the Durable Goods Report for insight into future economic health. If the Durable Goods Report indicates an increase in economic strength, the analysts could make adjustments to compensate for lower retention rates during these periods. Future research could be conducted to verify the relationship between this economic indicator and retention rates. If the relationship holds, then forecasting techniques might be effective in extrapolating future New Orders values as a proxy for economic strength; however, the application may be limited to how far into the future the prediction can go as the behavior of the economy is notoriously volatile. It may only be useful for one-month adjustments.

VI. Conclusion

6.1 Introduction

Since the inception of the All-Volunteer Force in the 1970s, the military has been in an on-going campaign to maintain adequate manning levels for current mission needs as well as having the flexibility to surge for future potential operations. Within the Air Force, maintaining adequate manning levels in the rated officer corps is the key to air power. Unfortunately, this task is a complex one that requires not only the capability to predict future conflicts but also predict retention behavior of officers within critical career fields. The task of maintaining adequate manning levels in the rated career fields is complicated by the fact that training and experience is critical to the execution of their mission. There is no equivalent one-for-one exchange of a brand new rated officer for a rated officer with years of experience; therefore, adjusting accession programs to ensure a large influx of rated officers is not enough. Retention of trained, experienced rated officers is vital to a well-balanced rated career field. Modeling retention behavior and determining factors that influence retention can aid in tailoring programs to ensure the sustainment of adequate manning levels in the rated career fields.

Modeling personnel retention patterns and behavior has been a topic of interest in both the public and private sector. Many modeling techniques, from Markov models to simulation, have been employed in the pursuit of finding the best retention model. Currently, the Air Force is modeling personnel retention using a five-year history of average retention rate in each year group. They use these average retention rates along with an assumption that current manning levels will not change in the next 30 years to

model a sustainment curve that they fit to current manning levels. These sustainment lines in conjunction with the congressional budget for manpower are used to determine the accession rates and retention programs for each career field each year. While this has proved to be a sufficient method for retention modeling, this thesis proposes changing the current method to a more statistical-based method, employing survival analysis over a larger time window to create survival curves rather than sustainment lines and potentially incorporating economic and political factors into the model to assist in fine-tuning the results.

By using survival analysis, this study has found that six demographic factors and one economic indicator are statistically significant factors in modeling the retention behavior of rated officers. These factors are marital status, gender, source of commissioning, prior enlisted service, whether the officer had dependents, and the New Orders value from the Advance Durable Goods Report. The results of these models also prove equivalent to current products created and used by HAF/A1. HAF/A1's model is sufficient at applying past retention patterns to current manning levels; however, it lacks any incorporation of outside influence to retention behavior. As Bezdek [14] stated, manpower forecasting is not robust without incorporation of the four categories: demographic, social, political, and economic variables. The benefit of using these new survival analysis models is the inclusion of a greater breadth of historical data, the flexibility of the model to incorporate as many or as few explanatory variables as desired, its potential predictive capability when incorporating economic indicators, its application in tailoring retention and access programs based on covariate influence, etc. There are multiple applications, but most crucial could be the added fidelity that external (e.g.,

economic) factors have on retention patterns. Incorporating economic variables into the model could aid in the forecasting fidelity of future retention patterns.

6.2 Limitations of Work

Several issues needed to be addressed during this investigation into retention modeling. First, modeling personnel retention behavior should be focused on the latest retention trends. Retention behavior in the 80s or 90s is not comparable to contemporary retention behavior, so this necessitated an analysis of the past ten years, which unfortunately and unavoidably contained at least three Force Management programs. These Force Management programs resulted in involuntary separations that may have skewed the attrition rates that were used in this study to model retention behavior. Additionally, this study relied on data that came from the MilPDS system. While this system is a powerful, effective tool in the storage and maintenance of personnel data, it is also prone to minor errors, including user input errors, system glitches, and missing data. Errors are traditionally overcome within this system by capturing monthly backups of the data that can be used for comparison and update. The system is not perfect, so a few minor errors in the system may still exist. For the purposes of this study, these errors were assumed to be scarce and insignificant. Finally, retention behavior among rated officers changes based on the airframe that they are associated with. Cross-training from one airframe to another could result in a contented rated officer with the intent to retain for 20 years to want to leave at the soonest opportunity. This continual shift in retention attitude based on voluntary and involuntary cross-training events makes modeling retention behavior in rated officers more complex than with non-rated officers who enjoy a more static professional environment.

6.3 Follow-On Research

Future studies into retention behavior of rated officers could include modeling retention within the pilot career field. Pilot retention is impacted by the specific airframe that the pilot flies. This is due to a change in several operational conditions including unit culture, unit climate, ops tempo, and mission focus. Specifically, fighter pilots and RPA pilots see the highest attrition rates and deserve an investigation into the factors that affect the desire to stay for 20 years.

Furthermore, this study only scratched the surface of the potential information that could be gleaned from using leading indicators such as the Advance Durable Goods Report to predict future economic strength. The New Orders value from the Durable Goods Report could be examined in more detail to see the extent of its potential to assist in predicting economic strength. Forecasting techniques could be applied to determine the level of predictive fidelity that may exist in this economic indicator.

Appendix A. SAS Code for Logistic Regression

```
/* Append all loss month for each year by serial number, SSAN
DATA "<insert filepath for new, appended file>";
    SET "<insert filepaths for each current loss month to be appended>";
    BY SSAN;
RUN;
/* Append all loss year files to one file by serial number, SSAN
DATA "<insert filepath for new, appended file>";
    SET "<insert filepaths for each current loss year to be appended>";
    BY SSAN;
RUN;
/* Append all last month inv files for each year by serial number, SSAN
DATA "<insert filepath for new, appended file>";
    SET "<insert filepaths for each current inv year to be appended>";
    BY SSAN;
RUN;
/* Create "retain" variable for each year file (Retain=1, Loss=0)
DATA "<insert filepath for new loss file>";
    SET "<insert filepath for current loss file>";
    RETAIN = 0;
RUN;
DATA "<insert filepath for new inventory file>";
    SET "<insert filepath for current inventory file>";
    RETAIN = 1;
RUN;
/* Append inv and loss by serial number, SSAN
DATA "<insert filepath for new, appended file with all records>";
    SET "<insert filepaths for inv and loss files to be appended>";
    BY SSAN;
RUN;
/* Check for duplicates
DATA "<insert filepath for file that lists all duplicate records>"
"<insert filepath for file that lists records without duplicates>";
    SET "<insert filepath for current file that may have duplicates>";
    BY SSAN ;
    IF first.SSAN and last.SSAN THEN OUTPUT "<insert filepath for file
that lists records without duplicates >" ;
    ELSE OUTPUT "<insert filepath for file that lists all duplicate
records>";
RUN;
PROC PRINT DATA = "<insert filepath for file that lists all duplicate
records>"
    TITLE 'RESULTS OF DUPLICATES DATASET';
RUN;
/* Delete duplicate records by serial number, SSAN; keep last year record
DATA "<insert filepath for new file without duplicates>";
    SET "<insert filepath for current file with duplicates>";
    BY SSAN;
    IF last.SSAN;
RUN;
```

/* Delete all records of non-rated officers

```
DATA "<insert filepath for new file with rated only>";
SET "<insert filepath for current file with non-rated included>";
IF off_cat ne 'Rated' then DELETE;
RUN;
```

/* Delete all rated non-flying officers, mis-categorized officers, and student rated

```
DATA "<insert filepath for new file with rated only/no students>";
SET "<insert filepath for current file with non-rated included>";
IF COREModel = '65F' then DELETE;
IF COREModel = '63A' then DELETE;
IF COREModel = '61S' then DELETE;
IF COREModel = '62E' then DELETE;
IF COREModel = '52R' then DELETE;
IF COREModel = '51J' then DELETE;
IF COREModel = '38F' then DELETE;
IF COREModel = '37F' then DELETE;
IF COREModel = '37A' then DELETE;
IF COREModel = '35P' then DELETE;
IF COREModel = '34M' then DELETE;
IF COREModel = '33S' then DELETE;
IF COREModel = '32E' then DELETE;
IF COREModel = '21R' then DELETE;
IF COREModel = '21A' then DELETE;
IF COREModel = '14N' then DELETE;
IF COREModel = '13S' then DELETE;
IF COREModel = '13D' then DELETE;
IF COREModel = '92T' then DELETE;
IF COREModel = '92M' then DELETE;
IF COREModel = '92J' then DELETE;
IF COREModel = 'UNK' then DELETE;
RUN;
```

/* Redefine variables of interest (DG, Dependents, Marital Status, and Prior Service

```
DATA "<insert filepath for new file with redefined variables>";
SET "<insert filepath for current file with old variables>";
IF Source_of_Commission = 'ACAD MIL SCI-ANG' then DG = 0;
IF Source_of_Commission = 'DG/HG AMS-ANG' then DG = 1;
IF Source_of_Commission = 'OTHACDG' then DG = 1;
IF Source_of_Commission = 'U.S.A.F. ACADEMY' then DG = 0;
IF Source_of_Commission = 'U.S.NAVAL ACADEMY' then DG = 0;
IF Source_of_Commission = 'US MILITARY ACAD' then DG = 0;
IF Source_of_Commission = 'AFACDDG' then DG = 1;
IF Source_of_Commission = 'DG ROTC 2-YR PGM' then DG = 1;
IF Source_of_Commission = 'DG ROTC 2-YR(FAG)' then DG = 1;
IF Source_of_Commission = 'DG ROTC 4-YR PGM' then DG = 1;
IF Source_of_Commission = 'DG ROTC 4-YR(FAG)' then DG = 1;
IF Source_of_Commission = 'ROTC 2-YR PGM-ANG' then DG = 0;
IF Source_of_Commission = 'ROTC 2-YR/FAG PGM' then DG = 0;
IF Source_of_Commission = 'ROTC 2-YR PROGRAM' then DG = 0;
IF Source_of_Commission = 'ROTC 4-YR PGM-ANG' then DG = 0;
IF Source_of_Commission = 'ROTC 4-YR PROGRAM' then DG = 0;
IF Source_of_Commission = 'ROTC 4-YR/FAG PGM' then DG = 0;
IF Source_of_Commission = 'DG OCS GRADUATE' then DG = 1;
IF Source_of_Commission = 'DG OTS GRADUATE' then DG = 1;
IF Source_of_Commission = 'OCS GRADUATE' then DG = 0;
```

```

IF Source_of_Commission = 'USAF OTS GRADUATE' then DG = 0;
IF DEPNTONR = '' then DEPENDS = 0;
IF DEPNTONR = '01' then DEPENDS = 1;
IF DEPNTONR = '02' then DEPENDS = 1;
IF DEPNTONR = '03' then DEPENDS = 1;
IF DEPNTONR = '04' then DEPENDS = 1;
IF DEPNTONR = '05' then DEPENDS = 1;
IF DEPNTONR = '06' then DEPENDS = 1;
IF DEPNTONR = '07' then DEPENDS = 1;
IF DEPNTONR = '08' then DEPENDS = 1;
IF DEPNTONR = '09' then DEPENDS = 1;
IF DEPNTONR ge '10' then DEPENDS = 1;
IF PS_EFY = '' then PRIORSVC = 0;
IF PS_EFY le '1' then PRIORSVC = 0;
IF PS_EFY ge '2' then PRIORSVC = 1;
IF MARITLST = 'A' then MARITALSTAT = 2;
IF MARITLST = 'D' then MARITALSTAT = 2;
IF MARITLST = 'I' then MARITALSTAT = 2;
IF MARITLST = 'L' then MARITALSTAT = 2;
IF MARITLST = 'M' then MARITALSTAT = 1;
IF MARITLST = 'S' then MARITALSTAT = 0;
IF MARITLST = 'W' then MARITALSTAT = 2;
IF MARITLST = 'Z' then MARITALSTAT = 0;

RUN;

/* Separate files by Pilot, CSO, ABM
DATA "<insert filepath for new file with ABMs only>";
  SET "<insert filepath for current file with all rated>";
  IF COREModel ne '13B' then DELETE;

RUN;

DATA "<insert filepath for new file with CSOs only>";
  SET "<insert filepath for current file with all rated>";
  IF COREModel = '11B' then DELETE;
  IF COREModel = '11E' then DELETE;
  IF COREModel = '11F' then DELETE;
  IF COREModel = '11G' then DELETE;
  IF COREModel = '11H' then DELETE;
  IF COREModel = '11K' then DELETE;
  IF COREModel = '11M' then DELETE;
  IF COREModel = '11R' then DELETE;
  IF COREModel = '11S' then DELETE;
  IF COREModel = '11T' then DELETE;
  IF COREModel = '11U' then DELETE;
  IF COREModel = '13B' then DELETE;

RUN;

DATA "<insert filepath for new file with Pilots only>";
  SET "<insert filepath for current file with all rated>";
  IF COREModel = '12A' then DELETE;
  IF COREModel = '12B' then DELETE;
  IF COREModel = '12E' then DELETE;
  IF COREModel = '12F' then DELETE;
  IF COREModel = '12G' then DELETE;
  IF COREModel = '12K' then DELETE;
  IF COREModel = '12M' then DELETE;
  IF COREModel = '12R' then DELETE;
  IF COREModel = '12S' then DELETE;

```

```

        IF COREModel = '12U' then DELETE;
        IF COREModel = '13B' then DELETE;
    RUN;
/* Separate by CYOS within each career field
DATA "<insert filepath for new file with all pilots>";
    SET "<insert filepath for current file with all pilots>";
        IF CYOS_EFY ge '20' then retain = '1';
        IF CYOS_EFY le '19' and retain='1' then DELETE;
RUN;
DATA "<insert filepath for new file with 0-6 CYOS>";
    SET "<insert filepath for current file with all CYOS>";
        IF CYOS_EFY ge '6' then retain = '1';
        IF CYOS_EFY le '5' and retain='1' then DELETE;
RUN;
DATA "<insert filepath for new file with 4-8 CYOS>";
    SET "<insert filepath for current file with all CYOS>";
        IF CYOS_EFY le '4' then DELETE;
        IF CYOS_EFY ge '8' then retain = '1';
        IF CYOS_EFY le '7' and retain='1' then DELETE;
RUN;
DATA "<insert filepath for new file with 8-14 CYOS>";
    SET "<insert filepath for current file with all CYOS>";
        IF CYOS_EFY le '8' then DELETE;
        IF CYOS_EFY ge '14' then retain = '1';
        IF CYOS_EFY le '13' and retain='1' then DELETE;
RUN;
DATA "<insert filepath for new file with 12-19 CYOS>";
    SET "<insert filepath for current file with all CYOS>";
        IF CYOS_EFY le '12' then DELETE;
        IF CYOS_EFY ge '19' then retain = '1';
        IF CYOS_EFY le '18' and retain='1' then DELETE;
RUN;
DATA "<insert filepath for new file with 20-22 CYOS>";
    SET "<insert filepath for current file with all CYOS>";
        IF CYOS_EFY le '19' then DELETE;
        IF CYOS_EFY ge '22' then retain = '1';
        IF CYOS_EFY le '21' and retain='1' then DELETE;
RUN;
/* Run logistic regression
PROC LOGISTIC DATA="<insert filepath for each CYOS file>" PLOTS=ALL;
    CLASS MARITALSTAT(PARAM=REF REF="0") SEX SOC PRIORSVC DG DEPENDS;
    MODEL RETAIN(EVENT='1') = MARITALSTAT SEX SOC PRIORSVC DG DEPENDS
/ CLODDS=PL;
RUN;

```

Appendix B. SAS Code for Survival Analysis

/* Get Survival Curves for all three Career Fields

```
PROC LIFETEST DATA="<insert filepath for current data>"(WHERE=(RETAIN=1))
maxtime=30 atrisk plots=survival(atrisk cb) outs="<insert filepath for
lifetest summary data output>";
    TIME CYOS_EFY*RETAIN(0);
RUN;
```

/* Get Survival Curves with strata comparisons

```
PROC LIFETEST DATA="<insert filepath for current data>"(WHERE=(RETAIN=1))
maxtime=30 atrisk plots=survival(atrisk cb) outs="<insert filepath for
lifetest summary data output>";
    STRATA SEX;
    TIME CYOS_EFY*RETAIN(0);
RUN;
```

/* Get Baseline Covariates (determine which variables stay in the model with PHREG)

```
ODS GRAPHICS ON;
PROC PHREG DATA="<insert filepath for current data>"(where=(retain=0))
PLOTS(TIMERANGE=(0,30))=SURVIVAL;
    CLASS MARITALSTAT SEX SOC PRIORSVC DG DEPENDS;
    MODEL CYOS_EFY*RETAIN(1) = MARITALSTAT SEX SOC PRIORSVC DG DEPENDS
/ SELECTION=STEPWISE SLENTRY=0.2 SLTAY=0.06 DETAILS;
    OUTPUT OUT=Outp XBETA=Xb RESMART=Mart RESDEV=Dev;
RUN;
ODS GRAPHICS OFF;
```

/* Plot Martingale and Deviance Residual Plots for Model Adequacy/Lack of Fit

```
TITLE "ABM Retention";
PROC SGPLOT DATA=Outp;
    YAXIS GRID;
    REFLINE 0 / AXIS=y;
    SCATTER Y=Mart X=Xb;
RUN;
PROC SGPLOT DATA=Outp;
    YAXIS GRID;
    REFLINE 0 / AXIS=y;
    SCATTER Y=Dev X=Xb;
RUN;
```

/* Load Covariate Settings (for Unique Survival Function Comparison)

```
DATA <insert desired name of covariate dataset>;
    INPUT SEX $ PRIORSVC DG DEPENDS MARITALSTAT SOC $;
    CARDS;
    F      0      0      0      0      1
    F      0      0      0      0      2
    F      0      0      0      0      3
```

F	0	0	0	0	4
F	0	0	0	1	1
F	0	0	0	1	2
F	0	0	0	1	3
F	0	0	0	1	4
F	0	0	0	2	1
F	0	0	0	2	2
F	0	0	0	2	3
F	0	0	0	2	4
F	0	0	1	0	1
F	0	0	1	0	2
F	0	0	1	0	3
F	0	0	1	0	4
F	0	0	1	1	1
F	0	0	1	1	2
F	0	0	1	1	3
F	0	0	1	1	4
F	0	0	1	2	1
F	0	0	1	2	2
F	0	0	1	2	3
F	0	0	1	2	4
F	0	1	0	0	1
F	0	1	0	0	2
F	0	1	0	0	3
F	0	1	0	0	4
F	0	1	0	1	1
F	0	1	0	1	2
F	0	1	0	1	3
F	0	1	0	1	4
F	0	1	0	2	1
F	0	1	0	2	2
F	0	1	0	2	3
F	0	1	0	2	4
F	0	1	1	0	1
F	0	1	1	0	2
F	0	1	1	0	3
F	0	1	1	0	4
F	0	1	1	1	1
F	0	1	1	1	2
F	0	1	1	1	3
F	0	1	1	1	4
F	0	1	1	2	1
F	0	1	1	2	2
F	0	1	1	2	3
F	0	1	1	2	4
F	1	0	0	0	1
F	1	0	0	0	2
F	1	0	0	0	3
F	1	0	0	0	4
F	1	0	0	1	1
F	1	0	0	1	2
F	1	0	0	1	3
F	1	0	0	1	4
F	1	0	0	2	1
F	1	0	0	2	2

F	1	0	0	2	3
F	1	0	0	2	4
F	1	0	1	0	1
F	1	0	1	0	2
F	1	0	1	0	3
F	1	0	1	0	4
F	1	0	1	1	1
F	1	0	1	1	2
F	1	0	1	1	3
F	1	0	1	1	4
F	1	0	1	2	1
F	1	0	1	2	2
F	1	0	1	2	3
F	1	0	1	2	4
F	1	1	0	0	1
F	1	1	0	0	2
F	1	1	0	0	3
F	1	1	0	0	4
F	1	1	0	1	1
F	1	1	0	1	2
F	1	1	0	1	3
F	1	1	0	1	4
F	1	1	0	2	1
F	1	1	0	2	2
F	1	1	0	2	3
F	1	1	0	2	4
F	1	1	1	0	1
F	1	1	1	0	2
F	1	1	1	0	3
F	1	1	1	0	4
F	1	1	1	1	1
F	1	1	1	1	2
F	1	1	1	1	3
F	1	1	1	1	4
F	1	1	1	2	1
F	1	1	1	2	2
F	1	1	1	2	3
F	1	1	1	2	4
M	0	0	0	0	1
M	0	0	0	0	2
M	0	0	0	0	3
M	0	0	0	0	4
M	0	0	0	1	1
M	0	0	0	1	2
M	0	0	0	1	3
M	0	0	0	1	4
M	0	0	0	2	1
M	0	0	0	2	2
M	0	0	0	2	3
M	0	0	0	2	4
M	0	0	1	0	1
M	0	0	1	0	2
M	0	0	1	0	3
M	0	0	1	0	4
M	0	0	1	1	1

M	0	0	1	1	2
M	0	0	1	1	3
M	0	0	1	1	4
M	0	0	1	2	1
M	0	0	1	2	2
M	0	0	1	2	3
M	0	0	1	2	4
M	0	1	0	0	1
M	0	1	0	0	2
M	0	1	0	0	3
M	0	1	0	0	4
M	0	1	0	1	1
M	0	1	0	1	2
M	0	1	0	1	3
M	0	1	0	1	4
M	0	1	0	2	1
M	0	1	0	2	2
M	0	1	0	2	3
M	0	1	0	2	4
M	0	1	1	0	1
M	0	1	1	0	2
M	0	1	1	0	3
M	0	1	1	0	4
M	0	1	1	1	1
M	0	1	1	1	2
M	0	1	1	1	3
M	0	1	1	1	4
M	0	1	1	2	1
M	0	1	1	2	2
M	0	1	1	2	3
M	0	1	1	2	4
M	1	0	0	0	1
M	1	0	0	0	2
M	1	0	0	0	3
M	1	0	0	0	4
M	1	0	0	1	1
M	1	0	0	1	2
M	1	0	0	1	3
M	1	0	0	1	4
M	1	0	0	2	1
M	1	0	0	2	2
M	1	0	0	2	3
M	1	0	0	2	4
M	1	0	1	0	1
M	1	0	1	0	2
M	1	0	1	0	3
M	1	0	1	0	4
M	1	0	1	1	1
M	1	0	1	1	2
M	1	0	1	1	3
M	1	0	1	1	4
M	1	0	1	2	1
M	1	0	1	2	2
M	1	0	1	2	3
M	1	0	1	2	4

```

M      1      1      0      0      1
M      1      1      0      0      2
M      1      1      0      0      3
M      1      1      0      0      4
M      1      1      0      1      1
M      1      1      0      1      2
M      1      1      0      1      3
M      1      1      0      1      4
M      1      1      0      2      1
M      1      1      0      2      2
M      1      1      0      2      3
M      1      1      0      2      4
M      1      1      1      0      1
M      1      1      1      0      2
M      1      1      1      0      3
M      1      1      1      0      4
M      1      1      1      1      1
M      1      1      1      1      2
M      1      1      1      1      3
M      1      1      1      1      4
M      1      1      1      2      1
M      1      1      1      2      2
M      1      1      1      2      3
M      1      1      1      2      4
;

RUN;

/* Run PHREG Regression on Baseline Covariates (Illustrates Range)
ODS GRAPHICS ON;
PROC PHREG DATA="<insert filepath for current data>"(WHERE=(RETAIN=0))
PLOTS(OVERLAY)=(SURVIVAL);
    CLASS MARITALSTAT SEX PRIORSVC DG DEPENDS;
    MODEL CYOS_EFY*RETAIN(1) = MARITALSTAT SEX PRIORSVC DG DEPENDS;
    BASELINE COVARIATE=<insert desired name of covariate dataset>
OUT=<insert desired name for data summary output data> SURVIVAL=_all_;
RUN;
ODS GRAPHICS OFF;

/* Use Baseline Covariates for Best and Worst Retention (Based on Odds Ratios)
DATA COVS2;
    INPUT SEX $ PRIORSVC DG DEPENDS MARITALSTAT SOC $;
    CARDS;
M      1      0      0      1      4
F      0      1      1      2      2
;

RUN;

/* Run PHREG Regression on Best and Worse Baseline Covariates
ODS GRAPHICS ON;
PROC                                PHREG                                DATA="I:\My
Documents\Thesis\Data\CombinedData\abm"(WHERE=(RETAIN=0))
PLOTS(OVERLAY)=(SURVIVAL);
    CLASS SEX PRIORSVC DG DEPENDS MARITALSTAT SOC;
    MODEL CYOS_EFY*RETAIN(1) = SEX PRIORSVC DG DEPENDS MARITALSTAT SOC;

```

```
        BASELINE COVARIATES=covs2 OUT=abmPHREGresults;  
RUN;  
ODS GRAPHICS OFF;
```

Appendix C. Kaplan-Meier Curves

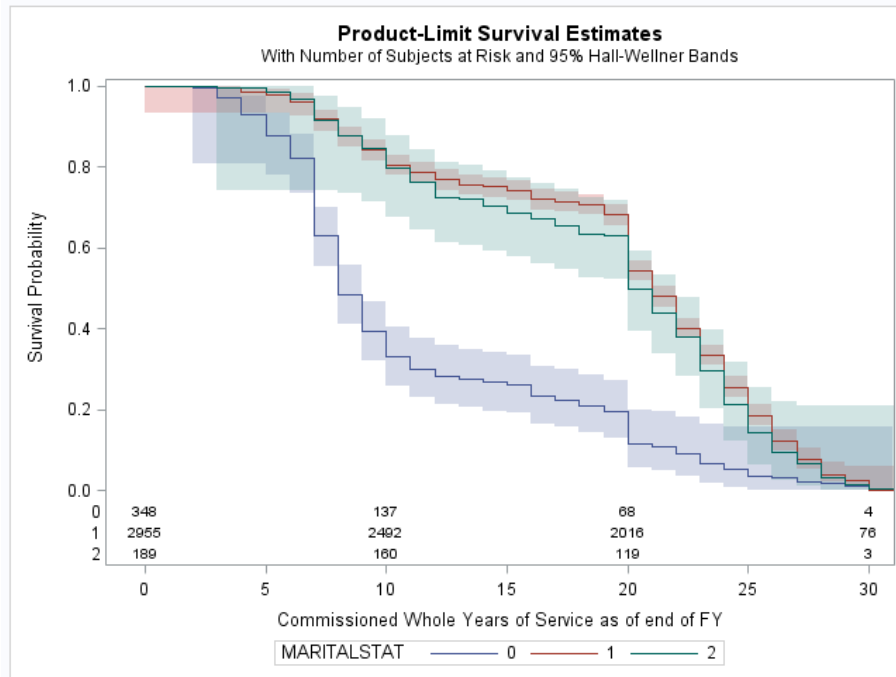


Figure 28. KM Curve for Marital Status (CSO)

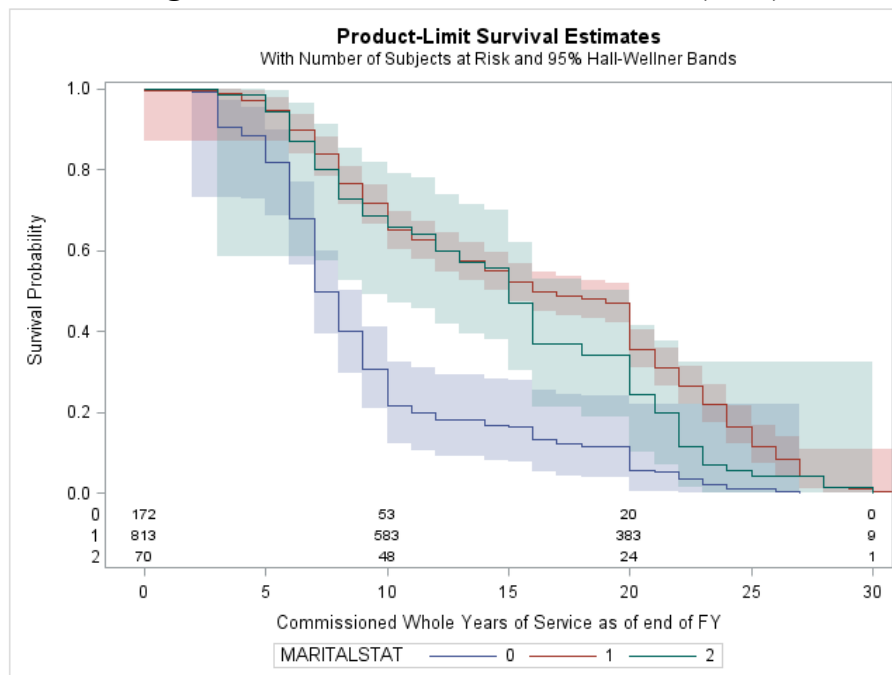


Figure 29. KM Curve for Marital Status (ABM)

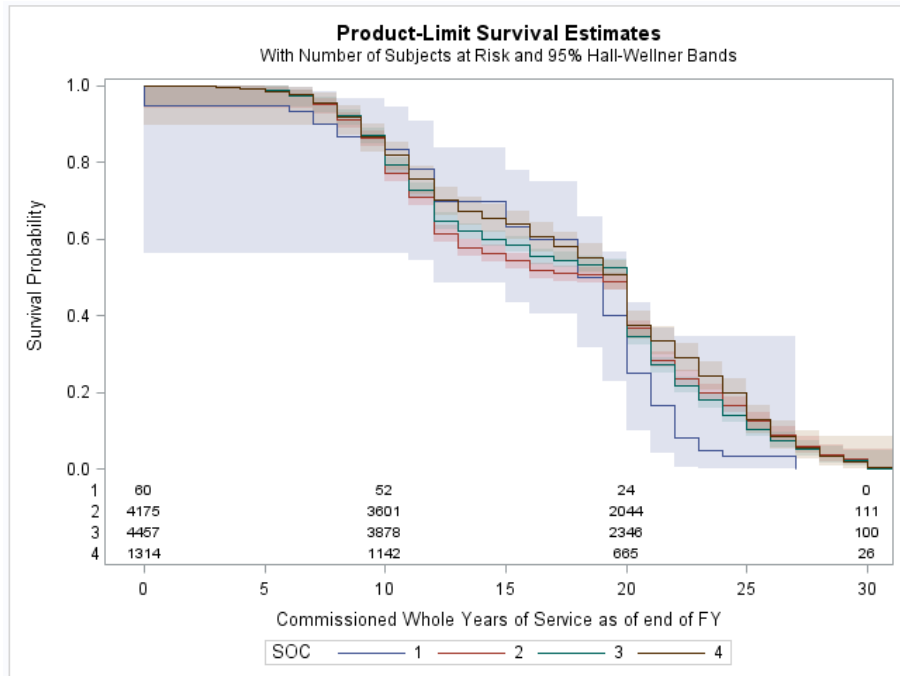


Figure 30. KM Curve for Commissioning Source (Pilot)

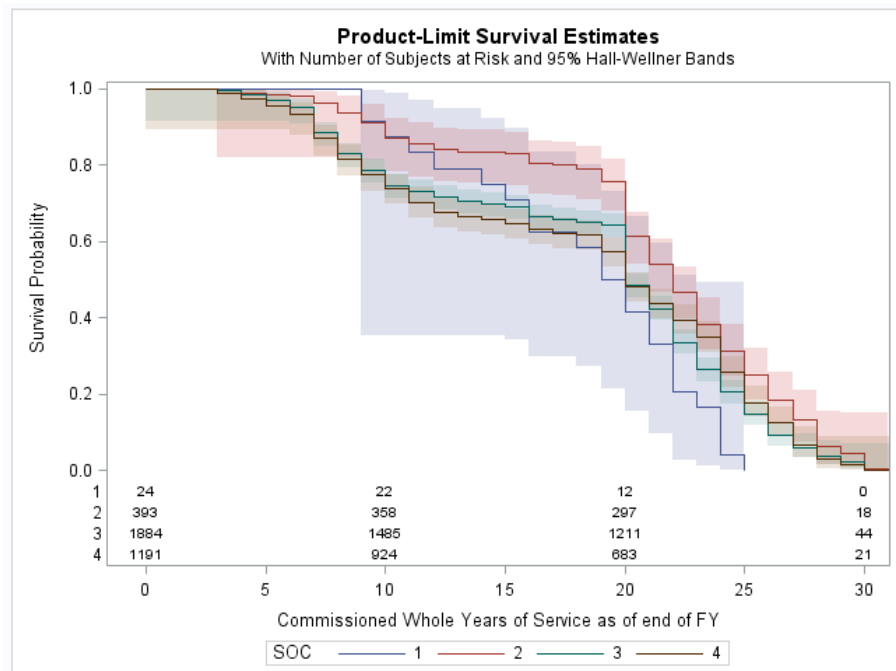


Figure 31. KM Curve for Commissioning Source (CSO)

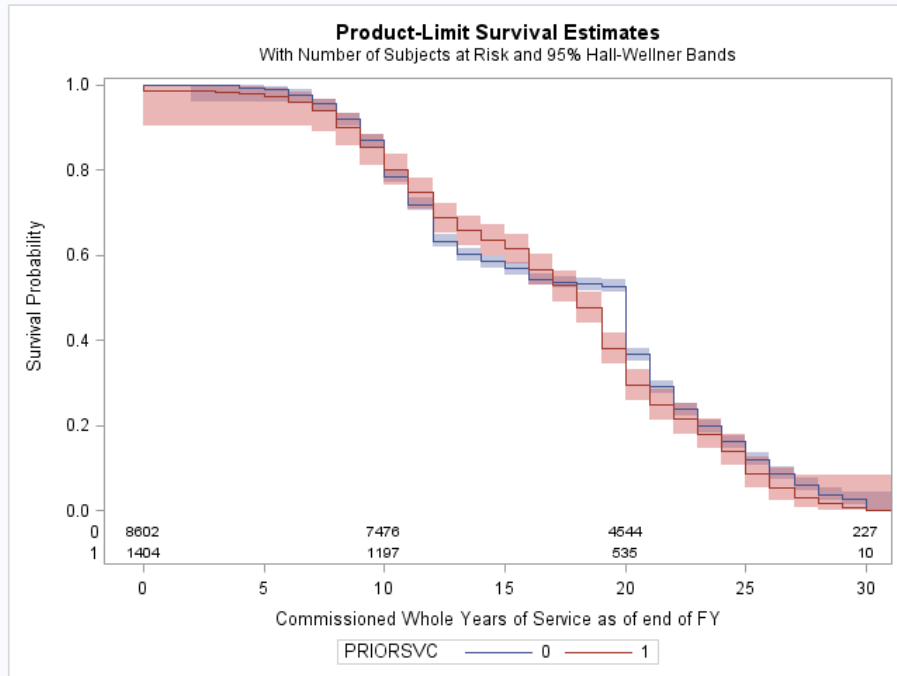


Figure 32. KM Curve for Prior Service (Pilot)

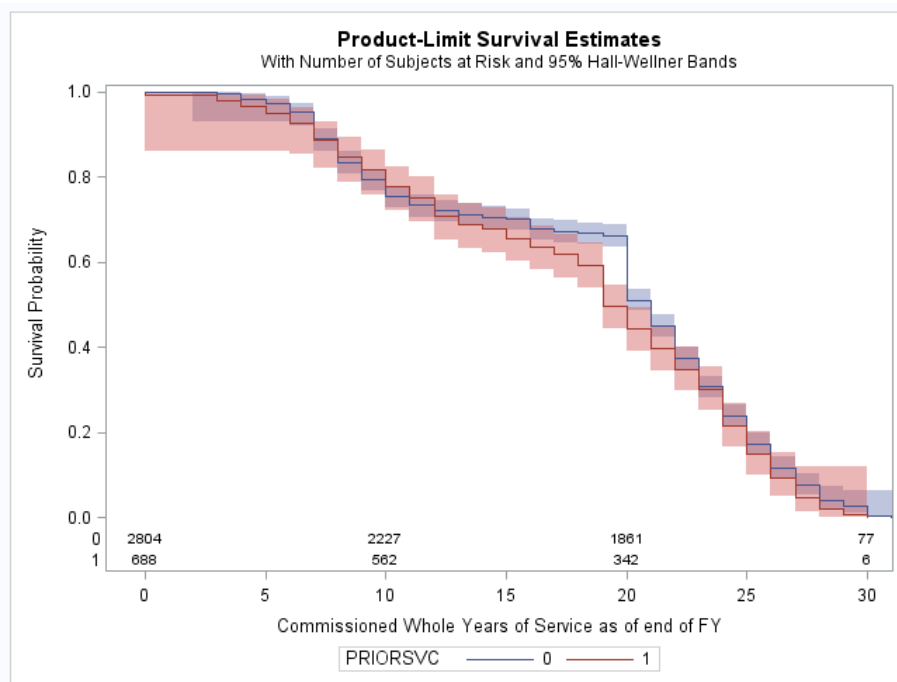


Figure 33. KM Curve for Prior Service (CSO)

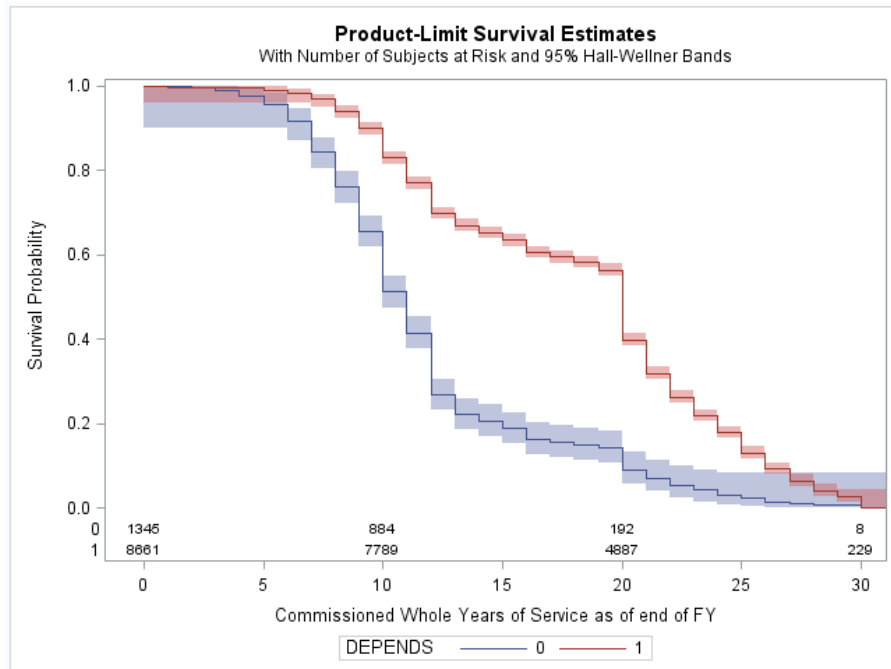


Figure 34. KM Curve for Dependents (Pilot)

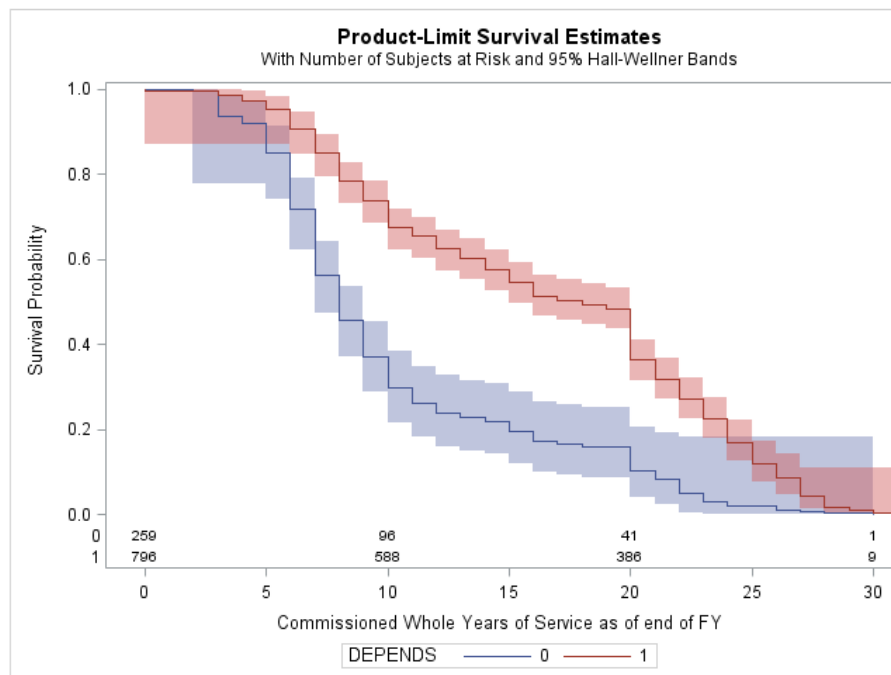


Figure 35. KM Curve for Dependents (ABM)

Bibliography

- [1] B. Asch and J. R. Hosek, "Looking to the Future: What does Transformation Mean for Military Manpower and Personnel Policy," RAND Corporation, Santa Monica, CA, 2004.
- [2] "Flying Operations: Aircrew Management," in *Air Force Instruction 11-412*, 2010.
- [3] United States General Accounting Office, "Military Personnel: Actions Needed to Better Define Pilot Requirements and Promote Retention," Washington, DC, 1999.
- [4] S. P. Barrows, "Air Force Pilot Retention: An Economic Analysis," Pennsylvania State University, Pennsylvania, PA, 1993.
- [5] "Personnel: Active Duty Service Commitments," in *Air Force Instruction 36-2107*, 2012.
- [6] CNN, "Air Force Faces Shortage of Drone and Jet Pilots," <http://www.cnn.com/2016/03/17/politics/air-force-pilot-shortage/>. Last accessed on 18 July 2016.
- [7] Air Force, "AFSC Fact Sheet," <http://www.af.mil/AboutUs/FactSheets/Display/tabid/224/Article/104484/officer-afsc-classifications.aspx>. Last accessed on 18 July 2016.
- [8] Congressional Budget Office, "Alternative Compensation Plans for Improving Retention of Air Force Pilots," Washington, DC, 1989.
- [9] "Statistical Analysis System," <http://support.sas.com/documentation/cdl/en/basess/58133/HTML/default/viewer.htm#a002645411.htm>. Last accessed on 30 July 2016.
- [10] "Microsoft Excel Support," <https://support.office.com/en-us/excel>. Last accessed on 18 July 2016.
- [11] J. Schofield, "Non-Rated Air Force Line Officer Attrition Rates Using Survival Analysis," Air Force Institute of Technology, Wright-Patterson AFB, OH, 2015.

- [12] C. Zens, "Application of Non-Rated Line Officer Attrition Levels and Career Field Stability," Air Force Institute of Technology, Wright-Patterson AFB, OH, 2016.
- [13] J. Capon and O. Chernyshenko, "Applicability of Civilian Retention Theory in the New Zealand Military," *New Zealand Journal of Psychology*, vol. 36, no. 1, pp. 50-56, March 2007.
- [14] R. Bezdek, Long-Range Forecasting of Manpower Requirements, New York: Institute of Electrical and Electronics Engineers, Inc., 1974.
- [15] R. Collins, S. Gass and E. Rosendahl, "The ASCAR Model for Evaluating Military Manpower Policy," Institute for Operations Research and Management Sciences, Maryland, 1983.
- [16] M. Tamiz, D. Jones and C. Romero, "Goal Programming for Decision Making: An Overview of the Current State-of-the-Art," *European Journal of Operations Research*, vol. 111, no. 3, pp. 569-581, 1998.
- [17] S. Gass, "Military Manpower Planning Models," *Computers Operations Research*, vol. 18, no. 1, pp. 65-73, 1991.
- [18] R. Hill, J. O. Miller and G. McIntyre, "Applications of Discrete Event Simulations Models to Military Problems.," in *2001 Winter Simulation Conference*, Piscataway, NJ, 2001.
- [19] J. Collofello, I. Rus, A. Chauhan, D. Houston and a. D. S.-D. D. Sycamore, "A System Dynamics Software Process Simulator for Staffing Policies Decision Support," in *Thirty-First Hawaii International Conference on System Sciences*, Kohala Coast, HI, 1998.
- [20] M. Gaupp, "Pilot Inventory Complex Adaptive System (PICAS): An Artificial Life Approach to Managing Pilot Retention," Air Force Institute of Technology, Wright-Patterson AFB, OH, 1999.
- [21] J. Hall, "An Analysis of Army Dentists Using Logistic Regression: A Discrete-Time Logit Model for Predicting Retention," Army-Baylor University Graduate Program in Health and Business Administration, Fort Sam Houston, TX, 2009.
- [22] J. Hall, "Forecasting Marine Corps Enlisted Attrition Through Parametric Modeling," Naval Postgraduate School, Monterey, CA, 2009.

- [23] Bureau of Labor Statistics, "Military Job Outlook," <http://www.bls.gov/ooh/military/military-careers.htm#tab-6>. Last accessed on 5 August 2016.
- [24] N. Sweeney, "Predicting Active Duty Air Force Pilot Attrition Given an Anticipated Increase in Major Airline Pilot Hiring," RAND Corporation, Santa Monica, CA, 2014.
- [25] Bureau of Labor Statistics, "Labor Force Statistics from Current Population Survey," <http://data.bls.gov/timeseries/LNS14000000>. Last accessed on 6 August 2016.
- [26] J. Kaplan, "What is Full Employment Rate of Unemployment," <http://www.colorado.edu/economics/courses/econ2020/section5/full-employment.html>. Last accessed on 6 August 2016.
- [27] Air Force Times, "Pilot Eligibility for \$225K Retention Bonus Expanded," <http://www.airforcetimes.com/story/military/benefits/pay/2015/04/06/pilot-eligibility-for-225k-retention-bonus-expanded-air-force/70822482/>. Last accessed 6 August 2016.
- [28] S. Chapman, "Keeping Pilots in the Cockpits," *AIR FORCE Magazine*, pp. 66-69, July 1997.
- [29] H. Jantscher, "An Examination of Economic Metrics as Indicators of Air Force Retention," Air Force Institute of Technology, Wright-Patterson AFB, OH, 2016.
- [30] Military Leadership Diversity Commission, "Officer Retention Rates Across the Services by Gender and Race/Ethnicity," no. 24, 2010.
- [31] G. Cerman and B. Kaya, "An Analysis of the Effect of Marital and Family Status on Retention, Promotion, and On-The-Job Productivity of Male Marine Corps Officers," Naval Postgraduate School, Monterey, CA, 2005.
- [32] M. Salas, "An Analysis of Promotion and Retention Factors Among Hispanic and Non-Hispanic Marine Corps Officers," Naval Postgraduate School, Monterey, CA, 2015.

- [33] A. Carter, "Person-Organization Fit and Its Effect on Army Officers with Less Than Eight Years of Active Duty Service," U.S. Army Command and General Staff College, Fort Leavenworth, KS, 2015.
- [34] D. Caswell, "USAF Female Pilot Turnover Influence: A Delphi Study of Work-Home Conflict," Air Force Institute of Technology, Wright-Patterson AFB, OH, 2016.
- [35] D. W. Hosmer, S. Lemeshow and R. X. Sturdivant, "Odds Ratio," in *Applied Logistic Regression*, 3rd ed., Hoboken, NJ, John Wiley & Sons, Inc., 2013, pp. 51-56.
- [36] UCLA: Statistical Consulting Group, "SAS Annotated Output Proc Logistic," http://statistics.ats.ucla.edu/stat/sas/output/sas_logit_output.htm. Last accessed on 28 January 2017.
- [37] RAND Corporation, "DOPMA/ROPMA Policy Reference Tool: Promotion Timing, Zones, and Opportunity," <http://dopma-ropma.rand.org/promotion-timing-zones-and-opportunity.html>. Last accessed on 28 January 2017.
- [38] J. Diaz, "Air Force Sustainment Lines," in *Air Force Institute of Technology: (OPER 501) Seminar*, Wright-Patterson AFB, OH, 2016.
- [39] UCLA: Statistical Consulting Group, "SAS Seminar: Introduction to Survival Analysis in SAS," http://www.ats.ucla.edu/stat/sas/seminars/sas_surival. Last accessed on 11 January 2017.
- [40] Diversified Technical Services, Inc., "Air Force Human Resources System Technical Support (MilPDS)," <http://www.dtsi.com/air-force>. Last accessed on 1 September 2016.
- [41] "Military Personnel Records System," in *Air Force Instruction 36-2608*, 2015.
- [42] "MilPDS Upgrade/Migration Guidance," <http://www.148fw.af.mil/shared/media/document/AFD-121011-037.pdf>. Last accessed on 1 September 2016.
- [43] A. Abel and B. Bernanke, "Business Cycle Facts," in *Macroeconomics*, 5th ed., New York, NY, Pearson Education, Inc., 2005, pp. 287-297.

- [44] M. Toosi, "Consumer Spending: An Engine for U.S. Job Growth," *BLS Monthly Labor Review*, pp. 12-22, November 2002.
- [45] U.S. Census, "Manufacturer's Shipments, Inventories, & Orders," http://www.census.gov/manufacturing/m3/historical_data/index.html. Last accessed on 21 November 2016.
- [46] Institute for Supply Management, "ISM Report on Business," <https://www.instituteforsupplymanagement.org/ismreport/mfgrob.cfm?SSO=1>. Last accessed on 21 November 2016.
- [47] Organisation for Economic Co-Operation and Development, "Consumer Confidence Index (CCI) (indicator)," <https://data.oecd.org/leadind/consumer-confidence-index-cci.htm>. Last accessed on 21 November 2016.
- [48] The Conference Board, "Consumer Confidence Survey," <https://www.conference-board.org/data/consumerconfidence.cfm>. Last accessed on 21 November 2016.
- [49] American Enterprise Institute, "Vital Statistics on Congress," <https://www.aei.org/publication/vital-statistics-on-congress/>. Last accessed on 21 November 2016.
- [50] "Air Force Strength Accounting Duty Status Program," in *Air Force Instruction 36-2134*, 2014.
- [51] W. Meeker and L. Escobar, "Censoring," in *Statistical Methods for Reliability Data*, Hoboken, NJ, John Wiley & Sons, Inc., 1998, pp. 3-41.
- [52] HAF/A1PF, "Career Field Health Charts," Air Force Personnel, 2016.
- [53] SAS Support, "Analysis of Residuals," https://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_phreg_sect052.htm. Last accessed on 11 January 2017.

REPORT DOCUMENTATION PAGE			<i>Form Approved</i> <i>OMB No. 0704-0188</i>	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.				
1. REPORT DATE (DD-MM-YYYY) 23-03-2017		2. REPORT TYPE Master's Thesis		3. DATES COVERED (From — To) Aug 2015 – Mar 2017
4. TITLE AND SUBTITLE Survival Analysis of US Air Force Officer Retention Rate			5a. CONTRACT NUMBER	
			5b. GRANT NUMBER	
			5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Franzen, Courtney N., Capt, USAF			5d. PROJECT NUMBER	
			5e. TASK NUMBER	
			5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology Graduate School of Engineering & Management (EN) 2950 Hobson Way WPAFB OH 45433-7765			8. PERFORMING ORGANIZATION REPORT NUMBER AFIT-ENS-MS-17-M-129	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) AF/A1PF Dr. Gerald Diaz 1500 W Perimeter Rd JB Andrews, MD 20762 gerald.diaz.civ@mail.mil			10. SPONSOR/MONITOR'S ACRONYM(S) AF/A1PF	
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Distribution Statement A. Approved for Public Release; Distribution Unlimited				
13. SUPPLEMENTARY NOTES This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.				
14. ABSTRACT As part of the effort to ensure proper retention rates for rated officers, retention models are created by the Air Force Personnel division that assist in predicting future retention patterns and accession needs. The techniques for creating these models, known as the "sustainment line," involve utilizing average retention percentages obtained from historical data. In this study, more statistical-based methods involving logistic regression analysis and survival analysis are utilized to obtain similar retention models for rated officers. The survival analysis curve produces similar results to the sustainment line, but the sustainment line currently employed is a one-dimensional view of retention patterns. It simply models the rate at which officers leave. The value of the survival curve created in this study is that it can be updated very quickly, is flexible in its construction, and can incorporate covariates into the model that are significant to retention rates. The Air Force has long known that there are external (e.g., economic) factors that impact retention. Using a survival analysis regression model instead of simply modeling the rate at which officers leave, this study was able to identify six demographic and one economic factor that may be significant to rated officer retention. This ultimately could lead to the creation of models that reflect the retention behavior of certain subtypes of officer and give insight that could be used to tailor retention and accession programs so that they are more resource-effective.				
15. SUBJECT TERMS personnel management, retention, regression, logistic regression, survival analysis, economic factors				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 109
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U		19a. NAME OF RESPONSIBLE PERSON Dr. Raymond R. Hill, AFIT/ENS
			19b. TELEPHONE NUMBER (Include Area Code) (937) 255-3636 x7469 raymond.hill@afit.edu	